PHD

Development and evaluation of resources for teaching evolution in primary schools

Buchan, Loredana

Award date:
2019

Awarding institution:
University of Bath

PHD

Development and evaluation of resources for teaching evolution in primary schools

Buchan, Loredana

Award date:
2019

Awarding institution:
University of Bath

Development and evaluation of resources for teaching evolution in primary schools

Loredana Buchan

A thesis submitted for the degree of Doctor of Philosophy
University of Bath
Department of Biology and Biochemistry
October 2018

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree. I am the author of this thesis, the work described therein was carried out by myself and interpreted together with my supervisor Laurence D. Hurst.

Table of Contents

Acknowledgements

I am extremely grateful to my supervisors, Laurence and Momna for their guidance and support throughout my study. Their insightful and encouraging comments helping me through some very challenging moments.

I would like to thank Liam, Rosina, Alex, Lucy and Joelle for their patience, expertise and computing skills which made the scripting and processing of my quantitative data possible. I would also like to thank Becky for making my PhD a little more manageable by sharing the insights she gained from her secondary school study and Marjorie for our regular chats over a coffee. Without the continued support of my family and friends I would not have been able to complete my PhD.

I am especially grateful to the teachers and students who participated in the study, without their help and support this research would not have been possible

Finally, must thank Jonathan who funded my PhD so generously allowing me the freedom of four years to complete this study. My visit to the Galapagos Islands, following in Darwin's footprints was a once in a lifetime experience for which I will be eternally grateful.

Summary

Teaching evolution in primary schools is important, not only to form the foundation of a planned spiral curriculum but also to address alternative conceptions before they have a chance to become entrenched. However, there is little research into how best to teach evolution in primary schools, with some doubting that abstract concepts of evolution can be understood in this age group. Whilst there are several untested proposed lessons, there is little evidence that evolution understanding in this age group can be assessed. This study presents evidence from two large-scale randomised trial tests collected over three consecutive academic years (tranche 1 2016/2017, tranche 2 2016/2017 and 2017/2018) using a validated and reliable assessment instrument adapted from the AAAS Project 2061 Evolution and Natural Selection test base. Different teaching schemes were developed, each of which led to significantly increased understanding in students of all abilities after correction for pre-teaching scores. Possibly of significance there is evidence of repeatable interactions between pairs of lessons, a phenomenon rarely considered in the literature. Additionally, the tranches of data show repeatable evidence for longer-term retention in both tranches with some waning in tranche 2. This study also sought to identify and confirm explanatory parameters at student, teacher and school level. Three student level predictors of improvement were examined after controlling for pre-test score using LOESS residuals, finding that student ability had a repeatable predictive ability, student gender was a weak predictor while age was not significant. Numerous class/school level predictors were considered by the study, with a teacher's self-reported perception of their increase in confidence level post teaching being the only repeatable predictor of student performance. Years of teaching experience, completion bias and the type of school attended were additionally found to be significant predictors in tranche 2. However, none of these teacher or school level predictors remained significant after multi test correction. In conclusion, this study suggests that all four of the teaching intervention programmes were effective in increasing student understanding of evolution with pairs of activities interacting in a positive and reciprocal manner.

Abbreviations

AAAS	American Association for the Advancement of Science
B.Ed	Bachelor of Education
BA	Bachelor of Arts
BSc	Bachelor of Science
CINS	Conceptual Inventory of Natural Selection
CofE	Church of England
CPD	Continued Professional Development
DNA	Deoxyribonucleic Acid
EAL	English as an Additional Language
ESI	Education, Skills and Training Deprivation Index
FSM	Free School Meals
GCSE	General Certificate of Secondary Education
ID	Intelligent Design
IDACI	Income Deprivation Affecting Children Index
IEA	International Association for the Evaluation of Educational Achievement
IMD	Index of Multiple Deprivation
KS2	Key Stage 2
LEI	Living Environment Deprivation Index
MATE	Measure of Acceptance of the Theory of Evolution
MCQ	Multiple Choice Question
MSE	Meeting Standard Expected
NC	National Curriculum
NOS	Nature of Science
NPD	National Pupil Database
NRC	Norwegian Refugee Council
Ofsted	Office for standards in education
PCK	Pedagogical Content Knowledge
SEN	Special Educational Needs
Scheme of Work	Scheme of Work
STEM	Science Technology Engineering and Maths
TIMSS	Trends in International Mathematics and Science Study

Chapter 1 Introduction

1.1 Chapter overview

In this chapter, the study and its rationale are introduced in context with the existing scientific and educational literature. The importance of early evolution instruction and how to teach it most effectively in primary schools is explored. Potential barriers to understanding evolution and the problems faced by teachers in the classroom are also discussed.

1.2 Introduction

There is growing recognition that young children benefit from studying evolution when biology is first introduced in primary school (Fail, 2008; Weiss & Dreesmann, 2014) when they are most receptive to new ideas and are actively questioning how the world works (Nadelson et al., 2009). Primary education helps to provide the foundation for evolution understanding (Akerson & Donnelly, 2010) and to develop a deeper understanding of evolution during progression through the school system (Wagler, 2012) as understanding of evolution tends to increase with greater exposure to it (Kim & Nehm, 2011). In 2014 the UK National Primary Science Curriculum was altered to include the conceptual understanding of Evolution and Inheritance as a statutory requirement for all year 6 students.

However, it is also recognized that evolution is an extremely problematic and widely misunderstood topic; see section 4.2.1 for a discussion of the main preconceptions held by children. As such, the introduction of this new content presents many challenges associated with teaching a complex science topic in primary schools, not least of which is the availability of age-appropriate resources for teaching evolution. Indeed, the bulk of the existing research focuses on the understanding of genetics and evolution in secondary school children, with very little known about the understanding of evolution in primary school children as reviewed by Venville and Donovan (2005) and Venville et al. (2005). The need to identify which teaching activities create the most effective learning experiences based on direct evidence has also been recognised by several authors (Beardsley, Bloom and Wise 2012; Glaze and Goldston 2015; Veal and Kubasko 2003), in no small part due to a lack of assessment tools which is particularly acute for primary-aged children (Ha et al., 2012; Nadelson & Sinatra, 2008; Nehm & Schonfeld, 2008; Wiles & Alters, 2011).

This project was designed to address three main areas of concern: poor teacher understanding of evolution, the lack of age appropriate resources and little guidance on how to teach the topic most effectively. An in-situ programme of teacher training was carried out to facilitate the standard delivery of one of four different randomly allocated Schemes of Work using resources developed and provided by the project. In order to assess the relative effectiveness of the intervention strategies within a quantitative framework a student assessment instrument adapted from the research of Flanagan and Roseman (2011) was developed for use with 9 to 11 year-old students.

1.3 Understanding concepts in evolution

1.3.1 Why is early evolution education so important?

According to philosopher Daniel Dennett evolution by natural selection is "the single best idea anyone has ever had" (Dennett, 1995, pp. 21-22) and it should be considered as ''the most powerful theory within the field of biology'' (Rutledge & Warden, 2000, p. 23). Eminent evolutionary biologists have asserted that the teaching of evolution is essential to a student's complete understanding of Biology (Dobzhansky, 1973; Gould, 1999).

There is growing recognition that young learners benefit from studying evolution when biology is first introduced in primary school (Fail, 2008; Kelemen et al., 2014; Keown, 1988; Pobiner, 2016; Wagler, 2010; Weiss & Dreesmann, 2014). It has been suggested that primary evolution education is needed to provide the foundation for the development of deeper evolutionary understanding further up the education system (McVaugh et al., 2011; Songer et al., 2009; Wagler, 2012) and that it is helpful to challenge intuitively constructed preconceptions (Williams, 2009). It has also been suggested that if these intuitive ideas are not addressed they can become entrenched and make it harder to acquire a scientifically accurate understanding of evolution (Järnefelt et al., 2015). Additionally, it has been proposed that early evolutionary education may even improve acceptance of evolution in adulthood (Beardsley et al., 2012; Lehrer & Schauble, 2004) and help to break the continued cycle of evolution education controversy (Hermann, 2011). In Wagler's opinion not teaching evolution at primary school level may negatively impact on a student's "overall long-term biological development" (Wagler, 2010, p. 444).

1.3.2 Are 9 to 11 year-old students able to understand evolution?

Whether primary age students are able to understand evolution needs careful consideration. In her case study on the acquisition of biological knowledge in children between the ages of 4-10, Carey (1985) proposed that between the ages of 8 and 10 do not have an adequate grasp of concepts such as inheritance, linkage and biological change necessary to understand evolution. This was supported by the research of Lawson and Thompson (1988) with slightly older American students (12 to 14) who suggested that some students are simply incapable of understanding natural selection because they have not yet developed the formal reasoning abilities needed to grasp it. While other research with primary age children has shown that understanding of science is linked to reasoning ability (BouJaoude et al., 2004; Cavallo, 1996; Tekkaya et al., 2012), the lower the reasoning ability of a child then the greater the number of alternative conceptions they hold and the harder it is for them to change their ideas after teaching (Lawson & Thompson, 1988; Oliva, 2003; Williams & Cavallo, 1995).Typically children are able to reason in evolutionary terms by the age of eight, reaching a transition phase between eight and nine when they start to understand microevolution in terms of intraspecific variation and confronting existential questions. Between the ages of ten and twelve (depending on family belief system) they are more willing to accept one kind of animal could be descended from a different kind – the beginning of a macroevolutionary understanding (Evans, 2008). More recent experimental studies have shown that children as young as 5 are able to grasp some ideas about genetics and natural selection if the correct type of instruction and scaffolding is provided (Evans et al., 2015; Kelemen et al., 2014; Legare et al., 2013; Metz, 2010; Shtulman et al., 2016; Venville & Donovan, 2007).

1.3.3 Why is evolution by natural selection so hard to understand?

Although Natural Selection has been described as a relatively simple concept by Mayr (1997) it has been shown to be very hard to understand (Ferrari & Chi, 1998; Gregory, 2009) and is far more difficult for students of all ages to grasp than most biologists imagine (Bishop & Anderson, 1990), due to the persistence of widespread alternative conceptions; see section 4.2. The difficulty of understanding evolution by natural selection is neatly summarized by Richard Dawkins in his book the Blind Watchmaker, “It is almost as if the human brain were specifically designed to misunderstand Darwinism, and to find it hard to believe” (Dawkins, 1986, p. 9). Accumulating evidence indicates that incorrect alternative conceptions of evolution and the process of natural selection are common and widespread throughout all ages and academic ability levels; secondary school students (Kampourakis & Zogza, 2008; Prinou et al., 2008; Settlage, 1994), undergraduates (Bishop & Anderson, 1990; Jensen & Finley, 1996) , student teachers (Asghar et al., 2007; Nehm & Schonfeld, 2007) and in-service teachers (Glaze & Goldston, 2015). For a partial review, see Gregory (2009).

There are a wide range of factors including: cognitive, pedagogical, epistemological, social, religious, affective and political constraints (Allmon, 2011; Nehm & Schonfeld, 2007; Pobiner, 2016; Smith, 2010a, 2010b; Thagard & Findlay, 2009) that contribute to the difficulty of understanding evolution and to an anti-evolutionary world view. Evolution is different from any other scientific theory as it challenges any idea that makes humans special and different from other animals, as well their perceived exception from the laws of nature (Richards, 2008). Some studies have suggested that evolution by natural selection should be classed as an emergent process (Chi, 2005; Petrosino et al., 2015; Wilensky & Novak, 2010) as it is based on an interconnecting network of fundamental disciplines including ecology, genetics and geology and so understanding evolution involves the assimilation and coordination of several different subjects and ways of thinking (Boggs et al., 2003). Emergent processes being notoriously hard for people to understand due to their high cognitive demand (Centola et al., 2000; Penner, 2000).

There is evidence that teachers cannot just add more knowledge to existing preconceptions to achieve a better understanding of evolution in their classes. Effective teaching of evolution is not just a matter of ‘bolting on’ new knowledge but it requires the ‘un-teaching’ and correction of alternative conceptions (Sinatra et al., 2008; Werth, 2012). Students are not ‘blank slates’ or ‘Tabula Rasa’ when they enter the classroom (Driver, Asoko, et al., 1994; Gilbert et al., 1982), they have formed their own ideas and theories to help them make sense of the world around them (Evans, 2008; Palmer, 1999). When students are taught science they interpret this new information and modify their own naïve theories (Wescott & Cunningham, 2005). Although this constructivist view of learning does have its limitations (Solomon, 1994), it is still the dominant model of science learning (Carey & Smith, 1993; Driver, 1989; Duit et al., 2008; Kuhn, 1993). Constructivist thinking assumes that learning is an active process in which the learner modifies their existing ideas to accommodate new information by generating rules and mental models that make sense of their experiences. Therefore to learn something new the students must change their mind (Colburn, 1994; Stebbins & Allen, 1975). Teachers should facilitate this process (Saunders, 1992) by using a variety of learning styles and approaches that provide the cognitive challenge needed to overcome their alternative conceptions.

Learning is the process of conceptual change in which the learner has to proactively revise and reorganize their existing knowledge (Wescott & Cunningham, 2005). This transition in student understanding from alternative conceptions to an accurate scientifically valid one is needed for some students to understand evolution by natural selection (Banet & Ayuso, 2003; Sinatra et al., 2008; Tanner & Allen, 2005). The main tenet of the conceptual change model is that students learn by assimilating and reconciling newly acquired information into what they already know (Cid, 2013) so that they undergo a holistic change (Posner et al., 1982). Conceptual change is extremely difficult to achieve as the accommodation of these new ideas requires that students must become dissatisfied with their existing conception, find the new conception plausible and understand why the new conception is better than their initial one.

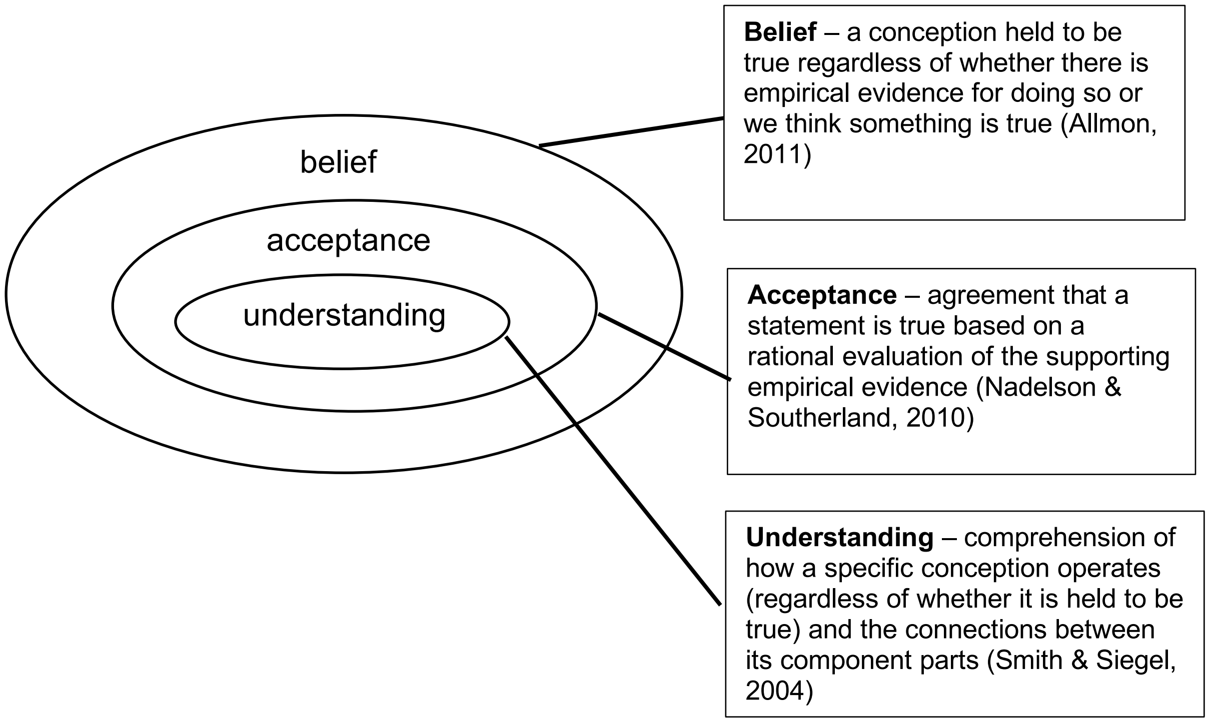
The language and terminology used to present and explain the evolutionary process and the understanding of its underlying concepts has been shown to be important (Werth, 2009) and should be a focus of attention (Evans, 2008; Kampourakis & Zogza, 2009). Words like need, selection, adapt and fitness all have everyday meanings which can be very confusing when learning about the subject (Bishop & Anderson, 1986; Gregory, 2009) especially if teachers switch back and forth between colloquial and the scientifically correct meanings (Nehm & Schonfeld, 2010; Rector et al., 2013; Wescott & Cunningham, 2005). Additionally, the use of biologically imprecise language to explain evolution can contribute to misunderstanding and reinforce misconceptions (Galli & Meinardi, 2011; Legare et al., 2013; Werth, 2013). For example the term selection can be confused with its every day meaning implying a conscious 'selector' that chooses something and which has a preference for a particular trait (Rector et al., 2013). This confusion is caused by implying that natural section is the cause rather than the mechanism of differential survival (Werth, 2012). Anthropomorphic language in which the forces of nature or evolution transforms an individual in a goal-directed way, is pervasive in educational settings (Legare et al., 2013) and should be avoided (Bishop & Anderson, 1990; Sinatra et al., 2008), as it reinforces alternative conceptions impeding the accurate understanding of biological processes and concepts (Legare et al., 2013).

Certain key evolutionary concepts have also been identified as being poorly understood in children; population thinking (Mayr, 1982), adaptation (Ohlsson, 1991), levels of organization (Mayr, 1997) randomness and probability (Mead & Scott, 2010b; Nadelson et al., 2009; Wilensky, 1993), the age of the earth (Jensen & Finley, 1996) and time frame involved in evolution (Samarapungavan & Wiers, 1997). Evolution is often seen as being progressive and purposeful acting towards the goal of improving each species (Mead & Scott, 2010a; Werth, 2012), some attributing evolutionary changes to intent or the agency of individual organisms (Garvin-Doxas & Klymkowsky, 2008; Kelemen & Rosset, 2009; Moore, 2002). Additionally, there are problems with the general understanding of science and how science works which negatively impact on the teaching of evolution (Coley & Tanner, 2012; Lombrozo et al., 2008; Wiles & Alters, 2011). Many conceive that science is based around a fixed set of facts or laws. Consequently, it may not be presented as a dynamic body of knowledge which is constantly reviewed and amended as new evidence emerges (Bennett et al., 1993; Driver et al., 1996). In scientific usage the word theory refers to "rigorously established and strongly supported ideas about the description of real world data that were obtained through observation, experimentation and analysis (Gunn, 2004, p. 9). Misunderstanding could arise from the common use of the term theory referring to ideas (Nadelson, 2009), imperfect facts (Gould, 1981) or guesses (Chuang, 2003). Chuang (2003) suggests that there is a general lack of understanding of the robustness of a scientific theory, with evolution being denigrated as 'only a theory' with little supporting evidence, when in fact its concepts are well grounded and based upon hard evidence (Dagher & Boujaoude, 2005; Nehm & Schonfeld, 2007). This lack of understanding about the nature of science (NOS) is one of the reasons students find learning about evolution so challenging (Chuang, 2003; Nehm & Schonfeld, 2007; Nelson, 2008).

Accumulating experimental evidence indicates a great number of other factors can also have an impact on the understanding of evolution. Children have been shown to worry about the ethical implications of the acceptance of evolution and its effect on society, fearing that it could lead to greater selfishness, reduced levels of spirituality, sense of purpose and self-determination (Brem et al., 2003). Woods and Scharmann (2001) posit that these social and emotional constraints regarding evolution may cause increased student misunderstanding of the theory. They go on to suggest that after religion the most frequently mentioned factor that shapes student attitudes toward evolution are personal relationships and the views of their parents, teachers and friends. Research conducted in the USA has shown that as children develop, their naïve biology beliefs are shaped by their exposure to different opinions about evolution (Cavallo & McCall, 2008) and become more similar to their parents' and common beliefs around them (Evans, 2001). Therefore in order to teach evolution effectively and successfully we need to understand and consider the world view of students in our classroom (Smith, 1994; Woods & Scharmann, 2001) and appreciate the differences between belief, acceptance and understanding.

Belief and acceptance are very different, even though many research papers treat these terms as interchangeable (Hermann, 2008). For the purposes of this thesis the term belief is defined as a person's subjective ways of knowing; their personal ideas about the world, using personal conviction, opinion and extra-rational criteria (Nehm et al., 2009; Sinatra et al., 2003; Smith, 1994). A person's belief in evolution is based on their convictions regardless of the evidence for or against it. Whilst acceptance is defined as the recognition of a concept's validity through rational and systematic evaluation of the evidence (Nehm et al., 2009; Sinatra et al., 2003; Smith, 1994); See Fig. 1.1. Science is not about belief but is about making inferences based on evidence, so to say a person accepts evolutionary theory means they think the theory accurately represents the understanding of the natural world based on critical evaluation of the evidence. Some authors advocate the avoidance of the term belief/believe within science classrooms to decrease the confusion between the scientific and religious uses of language (Padian, 2013; Werth, 2009).

Figure 1.1 schematic to show relationship between belief, acceptance and understanding, adapted from Allmon (2011, p. 651)



Should educators try to change their student's personal beliefs? Some authors maintain that the primary goal of science educators should be to change student beliefs particularly the non-scientific explanations for evolution by natural selection (Alters, 1997; Chinn & Samarapungavan, 2001; Lawson & Worsnop, 1992). Rice et al. (2015) go on to suggest that addressing creationist views may be the most effective way of improving evolution education but questioned the appropriateness of this approach. Other authors support the position that science educators should not set out to change their students' personal beliefs (Cavallo & McCall, 2008) as this is unethical (Meadows et al., 2000), morally questionable (Smith & Siegel, 2004), potentially infringes their religious beliefs (Southerland, Sinatra, et al., 2001) and makes it less likely that some students will learn about evolution (Kahan, 2015).

Some question whether student belief in evolution should be an educational goal (Scharmann, 2005) or even necessary, as several research studies have shown that disbelief in evolution is not a barrier to attaining an understanding of it (Hermann, 2013; Kahan, 2015; Lawson & Worsnop, 1992; Meadows et al., 2000) . Hermann (2008) proposed that rather than just asking students to believe something, science educators should encourage students to accept a concept based upon the relative merits of its supporting evidence. Therefore the role of a science educator should be to create an intellectually safe learning environment (Bramschreiber, 2014) in which the personal beliefs of their students should not be dismissed or underestimated when teaching evolution , shown by the small scale qualitative research of Hokayem and BouJaoude (2008). They should provide appropriate learning contexts and evidence (Nadelson & Hardy, 2015) to encourage students to examine, build upon and revise their own personal beliefs (Meadows et al., 2000; Nelson, 2012; Werth, 2012). This non-threatening approach has been shown to help students feel less alienated within the classroom (Hermann, 2013). For example the small scale research of Yasri and Mancy (2014) conducted with 17 to 18 year-old Thai students demonstrated that religious students could learn and accept evolution without it threatening their religious beliefs. The combination of religious belief with partial or full acceptance of evolution is common (Evans et al., 2010) and students frequently use coexistence models (Brem et al., 2003). Many prominent scientists advocate full acceptance with religious belief (Alexander, 2014; Collins, 2006; Gould & Haq, 1999; Miller, 2008) as do thousands of Christian clergy and hundreds of Jewish rabbis (Zimmerman, 2010). To avoid potential conflict educators must be careful not to attack deeply held personal and religious beliefs which lie outside the scope of science and should adopt a conciliatory tone (Lovely & Kondrick, 2008; Meadows et al., 2000).

There is also a complicated relationship between the acceptance and understanding of evolution, these two being identified as separate constructs (Cobern, 1994; Sinatra et al., 2003). It has been proposed that any real understanding of evolution cannot be achieved without acceptance of evolution and that lack of acceptance can be a serious barrier to understanding (Cobern, 1994; Davson-Galle, 2004; Lawson, 1983; McKeachie et al., 2002). Smith (2010a) concluded that religious beliefs especially fundamentalism decreased the acceptance of evolution and it has been shown to be a contextual barrier to its understanding in young Italian children aged between 7 and 9 (Berti et al., 2010). In contradiction, it has also been argued that students cannot accept a theory unless they develop at least some understanding of it (Lawson and Worsnop 1992, Sinatra et al. 2003). There is conflicting evidence to show that some students accept evolution without understanding it (Bishop & Anderson, 1990; Deniz & Donnelly, 2011; Lord & Marino, 1993; Sinatra et al., 2003; Wiles & Alters, 2011) and understand it without accepting it (Bishop & Anderson, 1990; Demastes et al., 1995; Lombrozo et al., 2008). Some studies have shown a strong positive correlation between these factors (Deniz et al., 2008; Rutledge & Mitchell, 2002) whilst others have shown none (Deniz & Donnelly, 2011; Rutledge & Sadler, 2011) or weak correlations (Mead et al., 2017). Likewise some studies suggest that acceptance can alter as a result of instruction (Lawson & Weser, 1990; Mead et al., 2017; Southerland, Sinatra, et al., 2001) whilst others do not (Bishop & Anderson, 1990; Lawson & Worsnop, 1992). Therefore, there is no clear relationship between these two factors or the direction of causality (Rissler et al., 2014; Smith, 2010a). It has been proposed that universal acceptance of evolution may not be possible in a typical classroom due to the diversity of world views; importantly, fundamentalist students who insist on the literal nature of their creation stories may never reconcile their religious beliefs with evolutionary theory (Cooper, 2001). Conflict with religious world view has been shown to be a major barrier to the acceptance of evolution in the US (Glaze, 2013; Meadows et al., 2000; Smith, 2010b). However, the acceptance levels of evolution in the UK is much higher than in the USA, with over 70% public acceptance compared to just a third (Miller et al., 2006), therefore low level acceptance of evolution is unlikely to be a major barrier to the understanding of evolution in UK primary schools. However, it is important to note that primary school teachers in the USA have been shown to have lower levels of acceptance than their secondary colleagues (Fowler & Meisels, 2010; Levesque & Guillaume, 2010; Losh & Nzekwe, 2011).

Perhaps a more appropriate goal to facilitate greater understanding of evolution should be to improve its acceptance through a better understanding of the Nature of Science (NOS) (Carter & Wiles, 2014; Lombrozo et al., 2008; McComas, 1998; Nadelson & Sinatra, 2010; Rudolph & Stewart, 1998; Smith & Scharmann, 1999) and general understanding of science (Nadelson & Southerland, 2010). The term Nature of Science can be described as the values and assumptions inherent to scientific knowledge and its development (Lederman, 1992) or "as a sort of shorthand for how science is done and what sorts of things scientists work on" (Reiss, 2008, p. 3). If students have a better understanding of Nature of Science then the teaching and learning of science including evolution should be made easier (Smith & Scharmann, 1999). Educators should aim to help students understand evolution by appreciating the tentative nature of science, analysing scientific evidence, practicing the processes involved and then evaluating the relative merit of the theory together with its component parts (Cavallo & McCall, 2008). Therefore giving students a better understanding of the modern Nature of Science should help them to form views that are more similar to those of the scientific community (Carter & Wiles, 2014) and facilitate greater acceptance of the theory of evolution.

1.4 Teaching and learning of evolution in primary schools

1.4.1 Evolution as a controversial topic

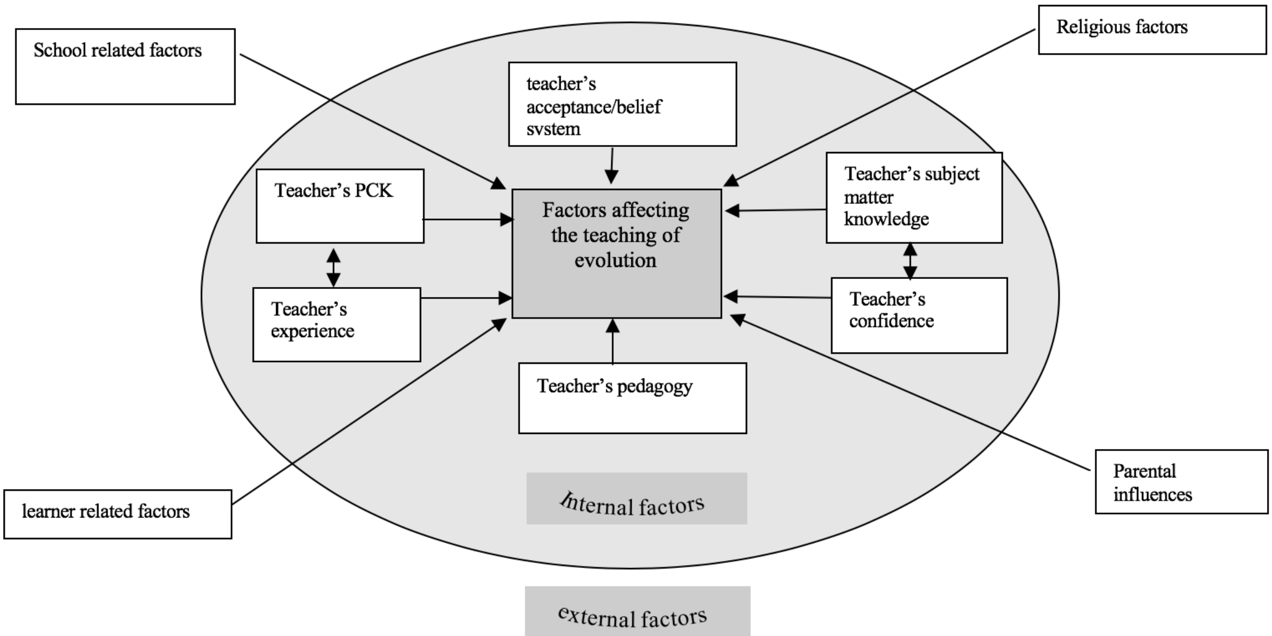
The teaching of evolution is still a controversial topic for some (Allgaier, 2009; Berkman & Plutzer, 2010; Bowman, 2008) especially in more secular Islamic countries like Turkey (Deniz et al., 2008) and the US where there is direct opposition to it being taught in schools (Berkman & Plutzer, 2011). 1 in 6 teachers in the USA is a Young Earth Creationist and 1 in 8 teach creationism as though it was a valid alternative to evolutionary theory (Berkman et al., 2008) with half of American Biology teachers advocating the teaching of creationism in their schools (Nehm et al., 2009). Religious beliefs are still one of the most persistent conflicts in the teaching and learning of evolution, especially regarding human creation and our common ancestry (Hokayem & BouJaoude, 2008; Rutledge & Mitchell, 2002; Trani, 2004). Religiousness has also been shown to be the strongest predictor of the rejection of evolutionary theory amongst the American public (Mazur, 2004). Allmon (2011, p. 655) outlined the apparent conflict between evolution and religion, “evolution appears to call into question the literal truth of some religious scripture". This conflict is not only present in the Judeo-Christian Bible but also the Quran and Hindu texts such as the Bhagavad Gita. It is commonly believed that Christian beliefs conflict with evolutionary theory (Scott, 2000) with learners themselves perceiving that religious beliefs are the major cause of conflict when learning about evolution especially in the USA (Woods & Scharmann, 2001). Studies have also shown that teachers can allow religious and other beliefs to influence what they teach (Nehm & Schonfeld, 2007; Trani, 2004).

Although the UK is more secular than the USA with a higher proportion of the population believing that religion is not an important part of everyday life, 31-44% in UK compared to 3-9% in the USA (Zuckerman, 2005), the antievolutionary and intelligent design movements are on the rise (Williams, 2008). Although creationism is predominantly a US protestant phenomenon (Martin, 2010) its influence is growing in the UK as reported by Reiss (2008). In a 2012 Angus Reid Public opinion poll 69% of British adults believed that humans had evolved from less advanced life forms whilst 17% were Young Earth Creationists believing that God had created human beings in their present form within the last 10,000 years. Curry (2009) found that 37% of primary and middle school teachers in England and Wales thought that creationism should be taught in schools alongside evolution. Whilst there is little empirical evidence on the extent of creationist teaching in UK schools (Williams, 2008) whether creationism has any place within the science classroom should be considered. The scientific community including leading scientists (Dawkins, 2006), Humanists UK (formerly the British Humanist Association), the National Secular Society, the Association for Science Education and the Royal Society (Royal Society, 2006) together with the UK government have rejected creationism and intelligent design (ID) and have rebuffed calls for its inclusion in the National Curriculum as a viable alternative to Darwinian evolution. They have stated that creationism cannot be used as an example of scientific controversy as there is no supporting evidence and there is a lack of acceptance within the scientific community. The UK government issued new guidance on the place of creationism and intelligent design in science lessons (Poole, 2008), banning the teaching of all pseudoscience and making evolution education mandatory in all state funded schools including free schools and academies. Educators should 'acknowledge the controversy' by explaining the conflict about evolution is cultural rather than scientific or do as Padian (2013) suggests, by teaching that it is not appropriate to present evolution as being controversial in the first instance.

1.4.2 What factors influence the teaching of evolution in the classroom?

There are a number of factors that can influence and affect how a teacher approaches the teaching of evolution in their classroom, both internal and external; see Figure 1.2. External factors are those that are imposed upon a teacher and which are out of their day to day control within the classroom. These include school related factors such as class size, location, type, and class organizational issues such as whether the classes are taught as single year groups or in mixed cohorts. These factors are all closely intertwined when teaching Key Stage 2 children in the UK, for example in the data collected for this thesis, children attending more rural schools tended to be taught in smaller primary schools in a single class mixed age cohort, whilst in more urban areas children attended either large primary or middle schools with multiple classes taught in single year cohorts. Each of these factors will now be considered separately.

Figure 1.2 Model of potential factors affecting the teaching of evolution in a class.



Note: Adapted from Figure 4: Model of factors affecting how teachers deal with potential controversy when teaching evolution (Sanders, 2010, p. 31)

External Factors

Class size

Several studies have reported a positive relationship between smaller class size and improved student performance in primary aged children (Finn & Achilles, 1990, 1999; Krueger, 1999; Krueger & Whitmore, 2001; Nye et al., 1999), particularly in class sizes of less than 20 and with at risk or minority students. Longitudinal studies have shown continued enhanced performance long after the students return to 'normal' sized classes (Finn & Achilles, 1999; Krueger & Whitmore, 2001; Nye et al., 1999). Brühwiler and Blatchford (2011) found that this positive effect remained significant even after accounting for confounding factors such as teacher qualifications and individual student characteristics and class context variables in upper primary school students. Rivkin et al. (2005) argued that it's not class size itself that improves student performance but the quality of teaching within the smaller class. It has been suggested that teachers of small classes like their students more, experience higher levels of morale and feel more satisfied with their own performance (Glass, 1982).They have a greater knowledge of their students and can differentiate learning experience for their individualized needs more easily (Finn & Achilles, 1999; Molnar et al., 1999). Students in smaller classes misbehave less so their teachers can spend less time on behavioural management (Betts & Shkolnik, 1999; Rice, 1999; Stasz & Stecher, 2000) and more time on planning and instruction.

Barr et al. (1983) suggest that improvements in student achievement can be enhanced by combining a reduction in class size with the adoption of different instructional techniques made possible by having less students in the class at any one time. Smaller classes allow more opportunities for assessment, feedback, writing, discussion, individualized help and less un-supervised time during group work whilst the teacher is occupied with another group (Betts & Shkolnik, 1999).

School size and type of school

The empirical evidence on school size and its relationship to academic performance is mixed and based from data collected in secondary schools. Some of these studies have reported either no significant effects (Deller & Rudnicki, 1993; Lamdin, 1995), significant negative relationships (Fowler Jr & Walberg, 1991; Heck & Mayor, 1993) or significant positive relationships (Sander, 1993). One reason for these discrepancies is that the relationship between secondary school size and academic performance has been shown to be a flattened inverted U shape, there being an optimum size for efficient functioning within the school beyond which no further increase in numbers provides further improvement in student outcomes (Bradley & Taylor, 1998). However, this pattern my not apply to much smaller primary schools, Mancebon and Molinero (2000) finding no relationship in the primary and junior schools they studied. However, school size is intimately linked with the pathway that different local education authorities follow through the compulsory years of schooling. There are 2 main pathways through the education system in the state schools encountered by this project: the more usual primary to secondary model and the junior to middle to college system; see Table 1.1.

Table 1.1 Summary of the different types of school in the UK

School type	Age range	Typical size of school	Typical classroom characteristics
Primary	5-11	Small	Single class teacher for all subjects Mixed ability and often mixed age group classes
Secondary	11-16 or 18	Large	Specialist teachers for all subjects Single year cohorts Streaming by ability

Junior	5-9	Small/medium	Single class teacher for all subjects Mixed ability and some mixed age group classes
Middle	9-13	Medium	Specialist teachers for science subjects usually in last two years Single year cohorts Streaming by ability or mixed ability classes
College	13-18	Large	Specialist teachers for all subjects Single year cohorts Streaming by ability

Note: Difference in shading indicates the two main educational pathways in the UK (Primary/Secondary or Junior/Middle/College)

Teachers in smaller schools tend to have to teach a wider range of subjects across the whole of the curriculum, restricting the opportunity for subject specialism and often have to undertake a wider range of administration tasks which potentially divert attention away from their teaching. However, due to their smaller size teachers in smaller schools may benefit from easier lines of communication making it easier to liaise with other teaching colleagues and their headteachers (Bridges & Hallinan, 1978). Whereas, in larger schools there is greater opportunity to stream students of similar ability creating more homogenous classes which are easier to teach and due to economies of scale they are able to invest in additional equipment like computers that enhance the learning experience for their students. There is also variation in the instructional approaches used in teaching primary and middle schools, Eccles et al. (1993) found that middle school teachers (of maths) taught in a more teacher centred way by controlling their students more and providing less opportunities for discussion, as opposed to the more student centred approach adopted in primary schools.

Religious affiliation of the school

Even if the teachers do not hold any particular strong religious beliefs themselves, religious factors such as whether the school is a faith school with a particular religious affiliation can still affect student progress. The Ofsted report of 2018 which analysed the National Pupil Database (NPD) of the Key Stage 2 2017 cohort found that students in faith schools made more progress in all subjects compared to students in schools with no designated religious character. This trend was supported by the research of Mancebon and

Molinero (2000) who found attending Church of England faith schools improved student performance.

Learner characteristics

There are a multitude of learner characteristics that can influence student achievement in science, such as, age, ability, special educational needs, whether English is the primary language, behavioural issues, socioeconomic status, ethnicity and their attitude towards learning.

Student’s attitude towards science

Martin et al. (2012) showed a positive and sustained correlation between student attitude towards science and science achievement, with students doing better if they enjoyed the subject in their analysis of the International Association for the Evaluation of Educational Achievement (IEA's) Trends in International Mathematics and Science Study (TIMSS) 2011 survey. Studies have shown a decline in attitudes towards science in the UK from age 11 onwards (Harvey & Edwards, 1980; Osborne et al., 2003) and there is some evidence that this decline begins in primary school (Jarvis et al., 2003; Murphy et al., 2007). Evidence from other countries suggests that students enter secondary school with positive attitudes and an interest in science which gradually erode as they experience more school science especially in female students (Kahle & Lakes, 1983).

Gender

The research of Breakwell and Beardsell (1992) with 11-14 year old British students showed that gender was the most significant variable in determining their attitude towards science. This result is supported by the literature review of Osborne et al. (2003) and meta-analyses of Becker (1989) and Weinburgh (1995) showing that boys in secondary school had a more positive attitude towards school science especially in physics. The research of Jovanovic and King (1998) with 12 year old American students reported that although there were equal levels of active participation and leading behaviours in the lessons they observed there was a decrease in the ability perceptions of female students. This perceptual difference was not factual as there was no significant difference between assessment levels due to gender. One explanation for this could be due to girls being more able to make comparative judgements as to their ability levels across the whole range of subjects (Eccles, 1985; Eccles et al., 1993) and so may perceive themselves as ‘better’ in other subjects compared to science.

Alternatively, students could be stereotyping science as a male domain (Eccles &

Blumenfeld, 1985) with implicit stereotypes regarding science participation and performance influencing their attitudes (Nosek et al., 2009). However, since the mid 1990's there has been a reversal in gender specific achievement with females doing as well if not better than males in most subjects at all levels of education (Elwood & Comber, 1995; Machin & McNally, 2005). From Bramley et al. (2015) analysis of 2014 GCSE results from the National Pupil Database (NPD) there is evidence that the gender gap in achievement had widened since 2004 with females outperforming males in all subjects but with a smaller gap for STEM subjects. However, there is a danger when analysing very large data sets that emphasis on statistical significance whilst ignoring the effect size can exaggerate the importance of sex differences (Hyde & Linn, 2006; Wilkinson, 1999). To date there does not seem to be a correlation between student gender and evolution understanding (Mead et al., 2017; Yates & Marek, 2014).

Other possible contributing factors

Data from the most recent TIMSS 2015 international survey showed that students qualifying for free school meals (FSM) which can be taken to be a measure of low socio-economic status in the UK, had lower achievement levels compared to their non-FSM peers. This pattern was confirmed internationally using 'books at home' as measure of social deprivation in other participating countries. Additionally, the English indices of deprivation published by the Ministry of Housing, Communities and Local Government (http://dclgapps.communities.gov.uk/imd/idmap.html) were used as a measure of the relative deprivation of the school's local catchment area.

Students with English as their first language were found to out-perform those with English as an additional language (EAL) as did Chinese students who out-performed all other ethnic groups. However, there is no evidence to suggest that ethnicity is related to evolution understanding (Woods & Scharmann, 2001). The age and ability of the students also need to be considered, as understanding is linked to reasoning ability and as children get older their ability to process abstract concepts has been shown to improve (Piaget & Cook, 1952). Conversely, the lower the reasoning ability of a child then the greater the number of alternative conceptions they hold and the harder it is for them to change their ideas after teaching (Lawson & Thompson, 1988).

Parental influences

There is evidence that some teachers fear confrontation with students and parents (Beard, 1996; Meadows et al., 2000; Sanders & Ngxola, 2009) when they prepare to teach evolution whilst other studies have shown that some do not want to teach the subject because of this continued conflict (Aguillard, 1998; Shankar & Skoog, 1993). To overcome this perceived conflict teachers might take the 'path of least resistance' (Goldston & Kyzer, 2009) by diluting the subject, avoiding it all together (Glaze & Goldston, 2015) or devoting minimal time to the topic in class (Berkman & Plutzer, 2011; Berkman et al., 2008; Sickel & Friedrichsen, 2013). On a more positive note Sanders (2010) found that although some teachers feared evolution before teaching it most of their concerns were alleviated once they had taught it for the first time as they received far fewer parental and student complaints than expected.

Internal Factors

Teacher's subject matter knowledge

Teachers are expected to be 'experts' but may lack the confidence and skills to be able to teach evolution effectively. Several research studies have reported that teachers do not feel adequately prepared to teach evolution (Aguillard, 1998; Griffith & Brem, 2004; Nadelson & Nadelson, 2010; Nadelson & Sinatra, 2010) with the commonest concern among teachers being related to inadequate knowledge (Sanders & Ngxola, 2009). According to the National School Workforce census (2013) 91.6 % of science lessons taught in secondary schools in the UK, were taught by teachers holding a relevant post A-level qualification. Unfortunately, there is no equivalent statistical information on primary school teachers, as to the nature of their highest qualification in biology or even if they hold a BSc, BA or B.Ed. The proportion of primary school teachers who hold a science degree will be significantly lower than their secondary colleagues and consequently it is quite common for primary school teachers to be educated up to GCSE or equivalent academic standard in biology (Nadelson & Nadelson, 2010; Papadopoulou et al., 2011). This lack of science knowledge causes to teachers to lack confidence in their ability to teach science (Sinclair et al., 2011) and as a consequence may reduce the time devoted to teaching in class, reducing the probability of their students meeting national science standards (Jesky-Smith, 2002). Additionally, primary school teachers in the USA have been shown to be less supportive of the theory of evolution than their secondary colleagues (Blank & Andersen, 1997).

There is no clear evidence to show that subject matter knowledge is related to teacher effectiveness and student achievement. Most studies report small, statistically insignificant relationships, both positive and negative (Andrews, 1980; Ashton & Crocker, 1987; Ayers & Qualls, 1979; Byrne, 1983; Haney et al., 1987; Quirk et al., 1973; Summers & Wolfe, 1975). The results may be mixed because subject matter knowledge is a positive influence up until some level of basic teaching competency in the subject but becomes less important thereafter, supported by the research of Monk (1994) who reported a curvilinear relationship between the improvement in student achievement and the number of related teaching courses completed by the teacher. However, teachers lacking adequate subject knowledge, especially in evolution are likely to have content misconceptions. They then may go onto teach these misconceptions to their students further compounding the problem (Jarvis et al., 2003).

Teacher's pedagogy and pedagogical content knowledge

Pajares (1992) found that a teacher's own schooling experiences and their years in education influenced their teaching pedagogy. The combined influence of these factors encouraging teachers to teach how they were taught (Deemer, 2004) and focus their teaching on the subject content they had studied (Alters & Nelson, 2002). "How teachers are taught, the way they teach, what they teach, and the misconceptions they hold" (Nadelson, 2009, p. 492) all influence their delivery of evolution in the classroom. Learning to teach is a long process in which teachers can be classified as learners in their own right who continue to learn as they teach. Experienced teachers develop a complex knowledge base over time that includes content knowledge, pedagogical knowledge, pedagogical content knowledge (PCK) which forms their own personal teaching framework (Shulman, 1987). Content knowledge is a teacher's knowledge of the specific content that has to be taught to their students (Shulman, 1986) whereas pedagogical knowledge is knowledge of how to teach effectively; the principles and techniques of classroom and behaviour management including knowledge of teaching and learning methods (Wilson et al., 1987). Teachers also need to develop pedagogical content knowledge (PCK), a unique understanding of how to help their students understand specific subject matter. This enables teachers to transform content knowledge into a form that students can use and make sense of (Sanders et al., 1993). In order to achieve this, they need an understanding of what students find confusing or difficult and be able to accurately address these issues in a variety of different ways (Feiman-Nemser, 2001; Sanders et al., 1993; Shulman & Grossman, 1988).

Teaching experience

Expert teachers have a sense of what works in their classroom and how lessons should flow (Sanders et al., 1993), however, years of teaching experience do not necessarily lead to teaching expertise (Carter et al., 1987). Studies have shown a positive relationship between teacher effectiveness and their years of experience (Klitgaard & Hall, 1975; Murnane & Phillips, 1981), as teaching requires a complex set of skills many of which can only be acquired in situ as they learn to teach. However, this relationship is not always significant or entirely linear. These studies have shown that inexperienced teachers with less than 3 years in class are less effective than more senior ones, the benefits of experience tending to level off after about 5 years especially in non-collegial work places. This curvilinear trend may be due older teachers failing to carry out continued professional development (CPD), selection effects caused by the promotion of the most experienced teachers into more administrative roles within schools (Vance & Schlechty, 1982) or attrition effects caused by the disproportional loss of more able teachers from the profession leaving a 'pool' of less capable senior teachers (Murnane & Phillips, 1981; Vance & Schlechty, 1982). However, more experienced teachers can continue to improve their performance if they are in a school setting that promotes continued professional development and collaboration (Rosenholtz & Kyle, 1984). Similarly, well prepared inexperienced teachers can be as effective as their more senior colleagues (Andrew & Schwab, 1995; Denton & Peters, 1988).

Teacher misconceptions

Moore and Cotner (2009) found that student knowledge structure about evolution tended to mirror their teachers and so any misconceptions the teachers held were likely to be passed on in their classrooms (Yates & Marek, 2013, 2014). Whilst (Harlen, 1997; Harlen & Holroyd, 1997; Pardhan & Bano, 2001) found that the alternative conceptions held by many teachers were similar to those of children, some of them at odds with the correct scientific view. However, it has been shown that teachers' alternative conceptions can be corrected by carefully planned interventions (Nehm & Schonfeld, 2007) so providing primary school teachers with better training, resources and coping strategies should improve the standard of teaching (Glaze & Goldston, 2015; Griffith & Brem, 2004). Abrie (2010) proposed that teachers should be provided with instructional materials and relevant professional development experiences that would provide them with the information needed to deal with misconceptions (be it their own or their students') and any potential conflicts in their classroom. The use of teaching materials that specifically address known preconceptions when presenting new information have been shown to give students a better chance of learning new concepts (Hewson & Hewson, 1983; Trumper & Gorsky, 1993) and have been successful in shifting students' conceptions closer to the accepted scientific view (Basili & Sanford, 1991; Fetherstonhaugh & Treagust, 1992; Smith et al., 1993; Thijs, 1992).

1.5 What is the best way to teach evolution in 9 to 11 year-old students?

In the literature several practices have been identified that seem to improve the success of evolution education. Teachers should 'know' their students and their existing ideas (Allmon, 2011; Yasri & Mancy, 2014) which would help to make the 'un-teaching' of their naïve alternative conceptions easier (Lawson & Thompson, 1988). They should focus upon the language used in their classrooms, avoiding anthropomorphic terminology (Evans, 2008; Kampourakis & Zogza, 2009) and the reinforcement of the vernacular meanings of scientific terms (Cunningham & Wescott, 2009; Pobiner, 2016). The adoption of a narrative rather than an expository style (Browning & Hohenstein, 2013) and the use of story based resources have been shown to 'free' children from conceptual constraints (Gerrig, 1993) whilst promoting interaction and interpretation of meaning (Doyle & Carter, 2003). It has been suggested that connecting the topic to everyday experience and making it relevant (Hillis, 2007) is an important way of engaging student interest.

When students are interested and actively engaged in their lessons they are more likely to learn (Chi, 2009), as the degree of student engagement is correlated with the likelihood of conceptual change (Dole & Sinatra, 1998). It has been shown that traditional didactic teaching methods are less effective in helping students develop scientific understanding (Brumby, 1984) than active teaching methods and resources (Nehm & Reilly, 2007). It has also been suggested that 'hands on' lessons (Beard, 1996) which are student centred and fun (Fail, 2008) are more effective. It has been shown that the use of inquiry based lessons (Demastes et al., 1995), case studies (Williams et al., 2012), modelling (Passmore & Stewart, 2002), and suggested that quantitative investigations that require measuring, recording, organization, data transformation and interpretation (Fail, 2008) all positively contribute to successful teaching. In addition, teaching of the historical development of the current theory of evolution has shown to be beneficial (Evans, 2000; Hermann, 2011; Jensen & Finley, 1995) and is supported by cognitive development theory (Shtulman, 2006).

Active learning strategies have been shown to increase engagement and enjoyment of evolution lessons in Canadian undergraduate students (Frasier & Roderick, 2011) and consequently to be more effective in promoting student learning (Alters & Nelson, 2002; Udovic et al., 2002) with greater retention of knowledge (Armbruster et al., 2009; Walker et al., 2008) compared to more traditional lecture type instruction. Active learning can be defined as any instructional method that engages students in their own learning and it may be best to think of it as an approach rather than a method (Prince, 2004). It requires students to complete meaningful learning tasks and to think about what they are doing and why they are doing it (Bonwell & Eison, 1991). Its key requirements are student activity and engagement in the learning process which can be subdivided into collaborative, cooperative and problem-based learning strategies. In collaborative learning students work together in small groups towards a common goal focusing on student interaction and has been shown to enhance academic achievement and student attitudes (Prince, 2004). Cooperative learning involves the careful structuring small groups so that students are able to learn from each other in a mutually supportive environment (Cooper et al., 2003; Millis, 2002). In both of these strategies the emphasis is directed towards cooperation between students rather than competition and has been found to enhance student learning compared to individual work (Johnson et al., 1998;

Springer et al., 1999). Finally, there is problem-based learning in which relevant problems are introduced at the beginning of the lesson and used to provide the context and motivation for learning. All of these active learning strategies effectively shift the focus away from the teacher to the learner, so that the teachers act as facilitators, who promote learning by creating safe environments in which students can take responsibility for their own learning (Frasier & Roderick, 2011). Not only do students in active learning environments perform academically better than in traditional didactic classes but there is the added benefit of the development other desirable skills such as hypothesis testing, critical thinking and problem solving (Barrows & Tamblyn, 1980; Ebert-May et al., 1997; Herreid, 1998).

The way in which scientific investigations are presented and structured within the classroom also needs consideration. The term learning cycle can be described as a way of structuring an enquiry to lead students through a scientific investigation step by step whereby they are actively engaged in the construction of knowledge and is intimately linked with conceptual change theory. In the 3E model described by Marek and Cavallo (1997) derived from Piaget's model of mental functioning (Piaget & Cook, 1952) there are three recognised steps that form the scaffold for science enquiries: exploration, explanation and expansion. During the exploration phase groups of students gather and record data and then answer pertinent questions about the activity so that they begin to assimilate this concept and become disequilibrated. Students then enter the explanation or concept development phase through careful teacher intervention when they construct a better understanding of the core concept by accommodation or re-equilibration. This can be achieved by using class data to construct graphs, tables or summaries so that trends can be identified and new vocabulary introduced. The final stage in this model of learning cycle is the expansion or elaboration phase in which the students are encouraged to apply their newly learned concept to different situations giving them the opportunity to organize and integrate it with their existing understanding. Marek (2008) went on to propose two more stages engagement and evaluation that 'top and tail' the previous model to form the 5E model. Whilst not derived from Piaget's model both are a useful and essential part of learning within the classroom. The engagement phase is used to introduce the activity and motivate the learning with evaluation assessing what the students know, their engagement and experimental skills.

1.6 Why was a directive scheme of work and teacher training needed?

To fully understand the complex and unifying role evolution plays requires knowledge drawn from different areas of biology as well as other diverse disciplines such as geology, history and mathematics. However, such knowledge is often not accessible to the teachers involved in its transmission (Tidon & Lewontin, 2004). Effective instruction in evolution requires teachers who have a good understanding of evolution and a firm grasp of its unifying role within biology (Dempster & Hugo, 2006) to enable them to connect the different parts of the curriculum into a coherent scheme with evolution at its core. Whilst the development of scientific subject matter has been a required and important part of teacher training since the introduction of the National Curriculum, there are still many primary school teachers with little knowledge of science (Jarvis et al., 2003) or lacking the training on how to implement student centred evolution active learning in their classrooms. Therefore, it was important to address these problems in order for teachers to minimize their students' alternative conceptions and enable them to develop an effective pedagogy (Osborne & Simon, 1996).

As teachers determine the quality of instruction within their own classrooms, it is vital that they are able to make professionally responsible and appropriate curricular decisions (Rutledge & Mitchell, 2002). However, this is an area that some teachers need help with as they are unsure as to what is appropriate, safe and even legal to include in their lessons (Bloom & Weisberg, 2007; Hermann, 2013). Many teachers have also been shown to experience problems translating guidance from reform documents such as the National Curriculum into the classroom (Crawford, 1999; Marx et al., 1994; Tobin et al., 1990), making it harder for them to understand how to organise the topic, divide it up appropriately and frame it for teaching. They also need better training, to provide them with an awareness of the tools that will help them cope with the internal and external factors involved in teaching evolution (Dotger et al., 2010; Glaze & Goldston, 2015) together with appropriate support to teach it in a conceptually sound way (Bandoli, 2008; Goldston & Kyzer, 2009). Fewer primary school teachers are comfortable teaching evolution than secondary ones (Fowler & Meisels, 2010) whilst Murphy et al. (2007) found that around 50% of primary school teachers identified a lack of confidence and their ability to teach science as a cause for concern. Teachers in these studies called for a substantial increase in science Continued Professional Development targeted at the more challenging aspects of the science curriculum in order to boost confidence levels which is supported by the findings of the Wellcome Trust review (2017). This 'State of the Nation' report on primary science education in the UK revealed that only 37% of schools had an allocated budget for continued professional development with 30% reporting that they had not received any support for teaching science in the last year.

The research of Harlen and Holroyd (1997) revealed that Scottish primary school teachers expressed concerns about the planning of science lessons and frequently made requests for help with respect to equipment and materials for science experiments, as a single lesson of practical work requires several hours of research, collection and checking of equipment. These findings were confirmed in the qualitative feedback collected by this thesis in which the majority of teachers reported minimal technician time (if any) making science experimental work harder to organise and less likely to be carried out in primary and middle schools. Additionally, when teaching outside of their subject specialism or an unfamiliar topic some teachers may act more like novice teachers (Hall & Hord, 2006) during the interactive phase of the lesson (Sanders et al., 1993) directing their classes to ‘safer’ activities; being more likely to devote more time to teacher talk, use resources such as worksheets and limit the opportunity for students to ask questions (Livingston & Borko, 1989). To facilitate a greater conceptual understanding in their students teachers need to have an understanding of the subject matter they are required to teach together with appropriate pedagogical skills and specific Pedagogical Content Knowledge (PCK) related to those concepts (Hall & Hord, 2006).

Teachers must also be aware of the typical difficulties that their students are likely to experience when learning about evolution (Shulman, 1986) and the common preconceptions held prior to instruction in order to tailor the lessons around their students (Novak et al., 2005). Whilst teachers can’t eliminate all barriers to learning about evolution, they should be able to recognize them and help guide students across them by teaching the topic in a clear, scientifically accurate way (Werth, 2012). Morrison and Lederman (2003) went onto suggest that it is unrealistic to expect teachers to have a comprehensive knowledge of these common misconceptions and advocate that being given an understanding of a few of the most relevant preconceptions for any given topic would be extremely beneficial. A guided reintervention approach was used as this study’s method of encouraging conceptual change and greater understanding of the topic. In this approach the teachers addressed common alternative conceptions in their lessons through the resources and training provided, but did not ask students to explicitly consider their preconceptions (Geraedts & Boersma, 2006). The goal being for the students to understand the evidence supporting the different components of the evolutionary theory, identify any alternative conceptions and then ‘reinvent’ them correctly.

Teachers have a unique responsibility as they are the primary source of student information and understanding of the topics they teach. The way in which they communicate the topic affecting the way in which students judge the trustworthiness of the information when deciding whether or not to believe it (Bloom & Weisberg, 2007). Teachers can compromise their teaching of evolution by presenting it in a way that makes their students doubt the accuracy and validity of the subject matter. Children have been shown to prefer to learn from knowledgeable (Koenig et al., 2004; Taylor et al., 1991) confident informants (Sabbagh & Baldwin, 2001). Sanders (2010) found that some teachers who were sceptical about evolution somehow communicated this to their students through their teaching. This was supported by the editorial of Harris and Koenig (2006) who suggest that a child’s beliefs is correlated to the level of testimonial support that a topic was given by trusted sources, showing why it is so important for primary school teachers to present evolution in the most positive way possible. For example, several studies have reported teachers presenting evolution as ‘only a theory’ (Asghar et al., 2007; Berkman & Plutzer, 2015; Miller, 2008), whereby implying that it was less robust and inferior to other scientific knowledge and lacked sufficient evidence to be considered as ‘fact’. This lack of understanding as to what constitutes scientific evidence (McComas, 1998) and the misunderstanding of the tentative nature of science needed to be addressed.

Therefore, improving the standard of evolution teaching was an important component of this study, particularly as a teacher’s attitude towards a particular topic has been shown to influence and impact on their curricular and pedagogical decisions (Carlsen, 1991; Friedrichsen et al., 2011; Grossman, 1990; Hashweh, 2005). Multiple studies have shown the positive impact that effective teachers can have on the achievement of their students (Marzano, 2003; Sanders & Horn, 1994; Yates & Marek, 2014), as they provide greater opportunities to learn about and retain relevant evolutionary concepts. Conversely, whilst good teaching is directly related to student understanding of evolution, unsatisfactory teaching has been linked to the persistence of alternative conceptions (Alters & Nelson, 2002; Goldston & Kyzer, 2009; van Dijk & Kattmann, 2009). Poor understanding of the processes involved in evolution and the Nature of Science have also been linked to poor representation of the topic (Deniz et al., 2011; Rutledge & Warden, 2000). Whilst other studies have shown the importance of helping primary school teachers develop a more conceptual and constructivist approach to science topics (Asoko, 1996; Neale et al., 1990).

1.7 Study rationale

The complex web of interacting factors that affect the teaching and learning of evolution need to considered before suitable teaching interventions can be designed and then evaluated to assess whether they address the core challenges facing evolution education (Nehm & Schonfeld, 2007).

Currently there is a distinct lack of research into the understanding of evolution in primary age school children as identified by Venville and Donovan (2005), there being only one existing study conducted in the UK (Browning & Hohenstein, 2013). Most studies have been conducted in research laboratories with small select groups which do not reflect realistic settings as they offer optimal learning environments free from distractions (Fisher et al., 2014). Consequently, there is also very little quantitative evidence about how to actually achieve a greater understanding of the topic within a classroom setting (Beardsley et al., 2012).See

Appendix A for a summary of existing research-based studies with primary age children.

Even more problematic is the dearth of studies on how to assess the effectiveness of any given intervention. Although, several authors have proposed possible activities for teaching evolution, ranging from classification by homologous structures (Chanet & Lusignan, 2008), modelling genes and DNA (Venville & Donovan, 2007) and interpreting evolutionary trees (Ainsworth & Saffer, 2013),the need to identify which evolutionary educational strategies and activities create the most effective learning experiences has been identified by several researchers (Glaze & Goldston, 2015; Nehm & Schonfeld, 2007; Veal & Kubasko Jr, 2003); "teachers need to know which approaches have direct evidence to support their use" (Beardsley et al., 2012, p. 288).

The lack of tools for the assessment of both the acceptance and understanding of evolutionary change particularly for use in primary age children, has been identified in several review and research papers (Ha et al., 2012; Nadelson & Sinatra, 2008; Nehm & Schonfeld, 2008; Wiles & Alters, 2011). Given the importance of evolution and the persistence of alternative conceptions this shortage of assessment tools is problematic for researchers and educators alike. There are two main assessment tools for the understanding of evolution, the Bishop and Anderson essay test (Bishop & Anderson, 1990) and the Conceptual Inventory of Natural Selection (CINS), a multiple choice test (Anderson et al., 2002), both of which were developed for use with undergraduate university students. Most of the other extant instruments for the assessment of evolution understanding being partially based on the Bishop and Anderson essay test (Demastes et al., 1995; Nehm & Schonfeld, 2007, 2008; Settlage, 1994). Similarly, the Measure of Acceptance of the Theory of Evolution (MATE) developed for use with adults by Rutledge and Warden (1999) is the most commonly used assessment instrument for the acceptance of evolution (Smith, 2010b).

The majority of the extant assessment tools for use with young children from reception through primary school (aged 3 to 11) have been phenomenographic in nature, involving the empirical study of the different ways in which they think about evolution. For practical reasons they have been implemented with relatively small sample sizes as they are based on the results of semi-structured interviews. In each case the interviews were audio taped, transcribed verbatim and then carefully scored by 'experts' using predetermined rubrics. For example, Samarapungavan and Wiers (1997), Solomon (2002) and Berti et al. (2010). Only one large-scale quantitative study using an assessment instrument specifically designed for middle and high school children exist. This was developed by Flanagan and Roseman (2011) as part of the AAAS Project 2061 focusing on improving evolution understanding in American children. The set of national standards based multiple-choice questions they developed for use with children aged 11 to 18 form the basis of the assessment instrument used in this thesis.

This distinct lack of large scale, in-situ, quantifiable data into the best way to teach evolution to upper primary school level and which factors influence the learning have been the motivation for this project. The limited number of small-scale phenomenographic studies providing ideas for some of the educational resources that have been adapted and integrated into comprehensive packages for non-science specialist primary school teachers.

1.8 Aims and Objectives

Aims

This project was designed to address the following research questions

1. What are the most effective ways of increasing understanding of evolutionary concepts among 9 to 11 year olds?
2. What factors affect the teaching and learning of evolution in the classroom?
3. Does teacher confidence and experience influence the learning of evolution in their classroom?

Evaluation strategies

The main objective of this project was to design and evaluate the relative effectiveness of four different experimental teaching intervention packages or Schemes of Work.

Phase 1: Development of teaching resource packs

A range of suitable resources were developed to enrich the conceptual understanding of genetics and evolution among 9 to 11 year olds. Four different Schemes of Work (SoW) were developed allowing the effect of two pairs of two different activities to be compared and the investigation of any possible interactions between them. This study compared two activities based on natural selection in the Peppered moth which either involved the students 'hunting' paper moths or a power point presentation in order to assess whether student performance was better in student centred or teacher centred tasks. Two activities studying homology and common ancestry in two different contexts; various species of extinct fossil Trilobites or extant mammal pentadactyl limbs including humans were compared to see if the context of the activity was important. It should be noted that all of the resources were based on existing small-scale research studies.

As part of this development a range of age appropriate teaching resources were created and piloted in two local primary schools. Although the primary school teachers were non-specialists (Grace et al., 2008; Murphy et al., 2007) they were experienced classroom practitioners. As such their opinions and suggestions on improving the teaching resources and assessment instruments were valued and used to make modifications during the consultation process. This input from primary school teachers and improvement in 'teachability' ensured that the resources had widespread support from the primary community.

Phase 2: Evaluation of teaching resource packs

The effectiveness of the 4 different teaching intervention programmes in improving the understanding of evolution was investigated using two randomly allocated large-scale tranches of data. The data were spilt into two tranches for logistical reasons over 3 consecutive academic years (tranche 1 2016/2017; tranche 2 2016/2017 and 2017/2018) allowing comparison between two independent samples. The resource packs were designed to allow the effectiveness of each main activity to be assessed within the sequence of lessons. Children were tested on their ability to answer 15 multiple choice questions adapted from the large scale American study by Flanagan and Roseman (2011); see section 2.3. Student assessment was carried out at 3 separate time points; a pre-test (before instruction), a post-test (carried out as soon as possible after instruction) and then a retention test (3-6 months post instruction). Analysis of this data allowing us to assess both short and long-term retention of the key evolutionary concepts.

Phase 3: Addressed the second research question, "What factors affect the teaching and learning of evolution in the classroom".

In order to investigate this interesting but complicated question, school related data were analysed and participating teachers were asked to complete a questionnaire which included demographic data and their:

- acceptance of evolution using the MATE assessment instrument from Rutledge and Mitchell (2002).
- knowledge of natural selection using CINS assessment instrument from Anderson et al. (2002).

- perceived confidence levels in evolution and how they changed after teaching the topic.
- years of teaching experience
- religious beliefs

1.9 Contribution of this study to the field of primary evolution education

The value of this study to the field of evolution education research lies in four main aspects:

1. This study consists of two tranches of large-scale quantitative data, unlike most other studies which tend to be small scale and based on data obtained through interviews or other qualitative measures.

2. This study assesses the understanding of evolution in UK primary school teachers and their students, there being very little relevant existing research due to the recent introduction of this topic into the Key Stage 2 National Curriculum.

3. The teaching interventions were carried out in classrooms by the students' normal teachers rather than by academics in laboratory settings. Thus, the results of this study are directly relevant and transferable to being applied across the country.

4. Unlike existing studies, this study analyses the effectiveness of a sequence of lessons rather than single activities.

1.10 Chapter summary

This chapter has outlined the background behind this study, the problem being investigated and the rationale of the investigation. This study has been placed within the context of the existing scientific and educational Literature and has cited relevant supporting studies. The next chapter discusses the methodology employed by the study.

Chapter 2 Methodology

2.1 Chapter overview

This chapter outlines the methods employed to conduct randomised trial tests and collect large scale quantitative data in the form of questionnaires from participating students and their teachers together with qualitative exploration of the topic via interviews. The choice of assessment instruments and teaching resources are discussed and their development into different teaching intervention programmes is outlined. Ethical issues are also considered.

2.2 Choice of research methodology

As part of the design process, two research approaches, qualitative and quantitative, were considered when deciding on the methodology. While it has been claimed that these two approaches are incompatible or irreconcilable, the so called ‘incompatibility thesis’; see Denzin (2012), there is a growing movement to combine the best practices and benefits of mixing both approaches in research (Teddlie & Tashakkori, 2003).

Quantitative methods are concerned with the techniques associated with the gathering, analysing, interpretation and presentation of numerical data (Teddlie & Tashakkori, 2009, p. 5) whilst qualitative methods deal with narrative data (ibid. p6) and can provide a source of well-grounded, rich descriptions and explanations of processes occurring in a specific context. Within this project questionnaires were employed with both teachers and students to collect large scale quantitative information efficiently and quickly. Interviews were conducted with participating teachers and with small student focus groups after completion to collect their views and ideas regarding the projects teaching resources and their experiences of taking part in the study. Calderhead (1996) promoted the appropriateness of qualitative study for investigating teachers’ roles in education. He particularly argued that research into teachers’ beliefs, experiences, contributions and involvement in innovation demanded a qualitative approach.

Both qualitative and quantitative approaches are useful for gathering different types of data from different perspectives and are both equally as important as sources of information. For this reason this study employed a mixed method design defined as “research in which the investigator collects and analyses data, integrates findings and draws inferences using both qualitative and quantitative approaches or methods in a single study” (Tashakkori & Creswell, 2007, p. 4) to enable a better description and exploration of the research questions and offered a methodology that was both confirmatory and exploratory. Qualitative data obtained through discussion with participating teachers and their open-ended written remarks were used in a complimentary way to allow elaboration, enhancement and clarification of the results from the quantitative analyses, in order to increase the interpretability and meaningfulness of the questionnaires (Greene, 1987; Rossman & Wilson, 1985; Shortland & Mark, 1987). It also provided a more in depth understanding of the research questions from different perspectives, capitalizing on the strengths of the methods and perhaps even offsetting the weaknesses of each (Mpeta, 2014).

2.3 Pilot testing was carried out

The teaching materials and student assessment instrument developed by this study were trialled in two separate primary schools during the summer of 2015. The purpose of the trials was to assess whether the planned activities were age appropriate and logistically possible within a class room setting and that the method of assessment was valid before collecting quantitative data. In both schools, pre-instruction training was given to the normal classroom teacher who was observed by the principle researcher using the resources provided. Pilot school 1 (n=16) was issued with Scheme of Work 1 whilst pilot school 2 (n=30) was issued Scheme of Work 4 allowing all of the different activities contained within the study to be scrutinised. Detailed field notes were taken during the observations which included suggestions made by the pilot teachers to improve the different activities; see Appendix E. Where appropriate the resources and assessment instrument were adapted to enhance the learning experience for the students and ease of use by the teachers. Full statistical analyses of both sets of pre and post test results were also carried out to confirm the validity of the assessment method before being used to collect quantitative data in the two tranches.

2.4 Student questionnaire was adapted from existing assessment items

In order to collect large-scale quantitative data, a paper and pencil multiple choice assessment instrument was developed by selecting assessment items from the AAAS science assessment website based on the research of Flanagan and Roseman (2011) . Flanagan and Roseman's research provided a rich source of reliable and validated assessment items that could be adapted to be used to assess primary and middle school children in the UK. Over 4 years they developed and field-tested a standards-based multiple-choice instrument with 9,419 students across 43 states from grades 6-12 (11 to 18 years old) across a broad range of demographics. It tested students' ideas about natural selection, similarities and differences between organisms, the Earth's continually changing environment, and common descent. Individual items for inclusion in the questionnaire were chosen for their relevance to the NC and appropriate cognitive demand. Each of the 15 items had four alternative answers, common alternative conceptions acting as distractors. The selected items fell into 5 broad categories: common ancestry/homology, natural selection, variation, fossils/geological time and extinction, allowing the assessment of the whole concept.

In order that the assessment instrument could be used successfully in mixed age group primary classes the assessment items had to be adapted to reduce reading difficulty and cognitive load whilst maintaining consistent item identity. In their original form the 15 selected questions had a mean Flesch reading ease score of 63.21 and a Flesch-Kincaid reading grade level of 7.17 (Kincaid et al., 1975) . The recommended reading ease score being > 60 for an average reader while a Flesch–Kincaid reading grade age of 7.17 equates to students of 12 to 13 years old. In the adapted version of the questions the reading ease score rose to 70.04 with a reading grade level of 5.88 which is more appropriate for primary age students. This reduction in reading difficulty was achieved by reducing the length and complexity of sentences and by using diagrams and tables of comparison rather than large chunks of text. (See Fig.2.1). A full breakdown and comparison of readability levels for each assessment item can be found in Appendix B. To compare the original version of the full student questionnaire with the amended one see Appendix C1 and C2)

Figure 2.1 Comparison of original and adapted assessment item EN046004 Flanagan and Roseman (2011)

Fig. 2.1a Original version

Some organisms, such as a chimpanzee and a human, have many similarities. Others, such as a zebra and a worm, have fewer similarities. What is TRUE about the ancestors of these organisms?

(A) Chimpanzees and humans share a common ancestor with each other, but zebras and worms do not share a common ancestor with each other.

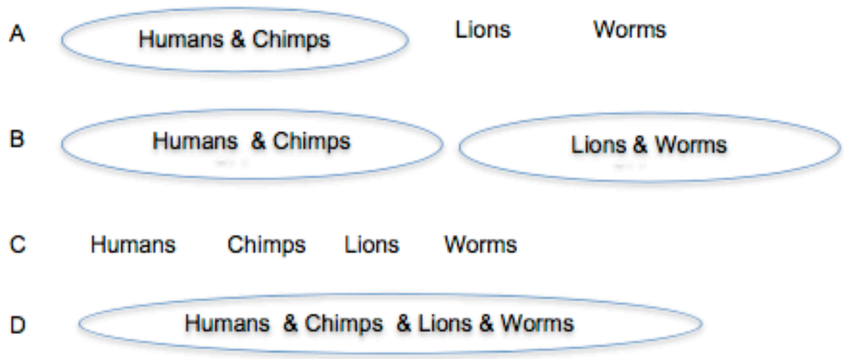
(B) Chimpanzees and humans share a common ancestor with each other, and zebras and worms share a common ancestor with each other, but chimpanzees and humans do not share a common ancestor with zebras and worms.

(C) Because chimpanzees, humans, zebras, and worms are separate species, none of them shares a common ancestor with any other.

(D) Chimpanzees, humans, zebras, and worms all share an ancient common ancestor.

Fig. 2.1(b) Adapted version of same assessment item (Lion substituted for zebra)

Some organisms like a human and a chimpanzee (chimp) have many similarities. Others like a lion and a worm have fewer. The **circles** show organisms that **share a common ancestor**. Which answer is **TRUE** about **who shares a common ancestor?** Select **one** answer from the options below**.**



During the pilot phase the full written amended version of the questionnaire (Appendix C2) was allocated to each student to read and complete individually. This format was found to take too long (~30-45 minutes) and concerns were raised about reading speed and concentration spans. At the pilot teachers' suggestion, the mode of delivery of the questionnaire was altered to one in which the teacher read out the questions whilst the students marked their responses on a grid as the test proceeded (Appendix C3 and C4). The students were given ample time to look at the question on the white board and think about their answers. To ensure consistency across schools the questions were read out in full and then summarized by focusing on the key differences in the alternative answers emboldened in the text (see Fig. 1(b)). This mode of delivery was much quicker to complete (~20minutes) and reduced problems associated with poor reading skills and had the added benefit of reducing missing data in assessment items appearing closer to the end of the questionnaire.

The students were assessed using the same test (to avoid adding an additional possible confounding factor) at 3 different time points: pre-teaching to establish their prior knowledge, as soon as possible after teaching (in practice around a week later in the next science lesson) to establish changes in understanding due to the teaching programme and if possible 3-6 months later to evaluate long term retention. Demographic data collected from the students was confined to name, gender and date of birth. In order to avoid problems associated with disclosing formal science attainment scores, teachers were asked to give their judgement of the relative science ability of each student within their classes as being either high (top 1/3 of class), middle or low (bottom 1/3 of class). To avoid requesting sensitive personal information, composite demographic data were taken from the school's most recent Ofsted report https://www.compare-school-performance.service.gov.uk/find-a-school-in-england and Indices of Deprivation explorer 2015 http://dclgapps.communities.gov.uk/imd/idmap.html for use in the school level analyses.

2.5 Teacher questionnaire was developed from existing assessment instruments

The pen and paper assessment instrument for teachers was designed to assess their acceptance and understanding of evolutionary theory and their perceived confidence level of teaching the topic. It also collected information on their length of teaching experience, religiousness and highest qualification in biology. The assessment instruments within this questionnaire were chosen for their strong convergent validity with the oral interview and ease of completion. This questionnaire was designed to take 15-20 minutes to complete before using the teaching intervention. This was to facilitate the collection of large-scale data. A shorter post teaching feedback sheet was also developed, to assess changes in confidence levels and any future needs.

The MATE assessment instrument was used to assess teacher acceptance of evolution. This is a 20-item questionnaire with built in contextual validity, based on a 5 point Likert-scale. It was designed to give a balance of simple positively and negatively stated items to ensure that it was relatively quick and easy to complete. The 20 items measured five different aspects related to acceptance of the theory of evolution (Rutledge & Warden, 1999). Participant acceptance was then scored from 20-100 possible points, with 20 being the lowest level of acceptance and 100 being the highest level of acceptance. The corresponding scores and categories for acceptance were 89-100, Very High Acceptance; 77-88, High Acceptance; 65-76 Moderate Acceptance; 53-64, Low Acceptance; and 20-52, Very Low Acceptance (Rutledge, 1996). The MATE instrument has reported Cronbach alpha coefficient reliability scores of 0.98 for secondary school biology teachers (Rutledge & Mitchell, 2002) and 0.94 for non-major biology undergraduates (Rutledge & Sadler, 2007). This implies that MATE instrument is a valid and reliable instrument to assess the acceptance of both these cohorts and by extension should be the most appropriate instrument to assess primary/middle school teachers as their levels of academic training/ability will probably fall somewhere in between these two groups. The MATE instrument will allow the exploration of the relationship between teachers' acceptance and their teaching of the subject.

To assess the understanding of evolutionary theory the CINS of Anderson et al. (2002) was used as it is based on actual scientific studies of natural selection. It consists of 20 multiple-choice items with common alternative conceptions as distractors. 10 key concepts were included in the CINS and each concept had a pair of items to explore its understanding. The test items had built in contextual validity as they were designed by subject area experts and addressed the 5 facts and 3 inferences described by Mayr (1992) as well as the underlying concepts of genetics and ecology that provide the foundation for natural selection. Each assessment item had an option of 4 responses.

In their research paper Nehm and Schonfeld (2008) compared the CINS to the ‘gold standard’ of oral interviews and validated the method an excellent replacement for the time consuming process of oral interviews, with a completion time of < 30 minutes. They found that the CINS gave rapid results, was easy and quick to score and interpret and therefore deemed the best for use with large samples. Although Anderson et al. did not develop this instrument to discriminate between participants, they suggested that a score of 16/20 or higher reflected a firm understanding of natural selection (Anderson et al., 2010).

In addition to the MATE and CINS assessment instruments the teacher questionnaire also collected relevant data to further explore aim 2 of the study. Perceived confidence levels in teaching science in general and each evolution strand of the NC were assessed on a 5point Likert scale.

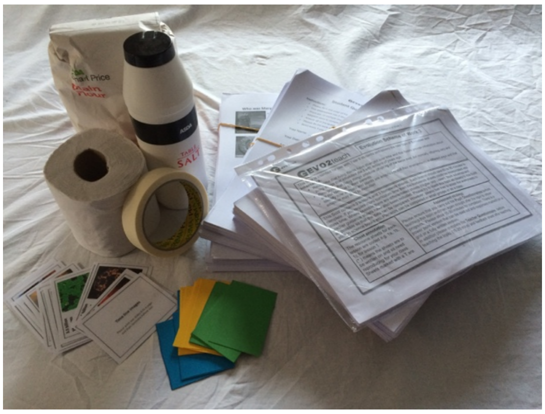
1	2	3	4	5
Not confident at all	Not confident	Fairly confident	Confident	Really confident

See Appendix D for full versions of the teacher questionnaires.

2.6 Choice of teaching resources was based on existing educational studies

Comprehensive detailed teaching Schemes of work were developed and adapted by liaison with partner schools. Improvements in ‘teacherbility’ ensured the resources were endorsed and used by primary school teachers. All resources developed for the project were fully differentiated and adjusted to be of the correct reading age (Kincaid et al., 1975) so that they were suitable for children of all abilities in mainstream schools. They were developed to cover the relevant parts of the Key Stage 2 National Curriculum and to support a scientifically valid understanding of evolution whilst avoiding emotional or religious conflict. The cost of the activities was minimized and used equipment that was easily available and suitable for use in the classroom. Participating schools received all materials so that there was no cost implication or additional preparation time.

Figure 2.2 Example of resource package provided for use in participating schools



Note: the full cost of the materials for each Scheme of Work was under £10 per class including printing and lamination.

Four different teaching intervention programmes or Schemes of Work (SoW) were developed to enrich the conceptual understanding of natural selection and evolution building upon pertinent biological concepts introduced earlier in their primary education. Students had previously learned about fossils in the year 3 rocks topic (describe in simple terms how fossils are formed when things that have lived are trapped within rock) and through the topic of living things and their habitats in years 4 and 5 had gained an awareness of the variety of living things, classification, and how living things are adapted to their different habitats and interact together. The teaching intervention programmes were designed to cover the following aspects of the newly revised Year 6 Key Stage 2 National Curriculum:

1. *Recognise that living things have changed over time and that fossils provide information about living things that inhabited the Earth millions of years ago*

2. *Recognise that living things produce offspring of the same kind, but normally offspring vary and are not identical to their parents.*

3. *Identify how animals and plants are adapted to suit their environment in different ways and that adaptation may lead to evolution.*

The lesson order: variation, natural selection/microevolution, geological time and then macroevolution, was kept the same to facilitate a greater understanding of evolution within a constructivist framework and to minimise the number of variables. This best practice restriction accords with recent randomized control trial indicating that, at least in secondary schools, teaching genetics before teaching evolution causes a marked improvement effect on evolution understanding, at no detriment to genetics understanding, compared with the opposite treatment sequence, (Mead et al., 2017). A lesson on geological time scales was included to improve understanding of macroevolution and to help the students contextualise the huge periods of time involved.

The lessons were structured according to the 3E learning cycle based upon Piaget's model of cognitive development (Piaget, 1964): Exploration, Explanation and Expansion; together with the enquiry-based learning method of Van de Walle (1990) the standard school lesson format used in schools. Each lesson consisted of three separate components: a starter activity to introduce the lesson and establish prior knowledge, a main or work phase activity and a plenary to consolidate. The starters and plenaries were novel or adapted from pre-existing teaching resources and were identical in each of the 4 teaching packages. This was to ensure the different work phase activities were embedded within the same conceptual framework and their relative impact could be evaluated during data analysis. The work phase activities for lesson 1 and lesson 3 were identical in each teaching package. There were 2 alternative work phase activities for lessons 2 and 4. This made a total of four different pathways through the teaching materials giving rise to the four different teaching packages. This arrangement also allowed the impact of the work phase activities of lessons 2 and 4 the only resources that were different to be evaluated separately and for possible interactions to be identified. See Table 2.1 for a summary of the different work phase activities.

Table 2.1 Schematic outlining the content of the work phase activities of the 4 different Schemes of Work (SoW)

Scheme of Work	Lesson 1	Lesson 2	Lesson 3	Lesson 4
1	Quantitative investigation of variation within the class	Investigating natural selection in Peppered moths. Student centred active hunting activity	Investigating geological time with the toilet roll time line	Investigation of homology and common ancestry in Trilobites.
2	Quantitative investigation of variation within the class	Investigating natural selection in Peppered moths. Student centred active hunting activity	Investigating geological time with the toilet roll time line	Investigation of homology and common ancestry using the pentadactyl limb.
3	Quantitative investigation of variation within the class	Investigating natural selection in Peppered moths. Teacher centred activity	Investigating geological time with the toilet roll time line	Investigation of homology and common ancestry in Trilobites.
4	Quantitative investigation of variation within the class	Investigating natural selection in Peppered moths. Teacher centred activity	Investigating geological time with the toilet roll time line	Investigation of homology and common ancestry using the pentadactyl limb.

Note: Lesson 1 and lesson 3 were the same in all Schemes of Work. There were two different main activities in lesson 2 and 4, giving a total of 4 different pathways for use in schools. Each colour representing a different main activity.

The work phase activities for lessons 1 and 3 were based on suggestions from other researchers in the field and were the same for each of the four Schemes of Work. The work phase activity of the first lesson (Fig. 2.3) of each teaching intervention was to introduce the existence of variation among individuals and to highlight intraspecific variation; a prerequisite for the correct mechanistic understanding of natural selection (Shtulman & Schulz, 2008). Intraspecific data were collected in class and then mathematically processed in order give a wider appreciation of the variability within a species and help to overcome essessentialist thinking (see section 4.2.1)

Figure 2.3 Photographs showing the main activity from lesson 1 which involved estimating and measuring handspans and then and transforming the data into bar charts as example of intraspecific variation within the class.

The scale of geological time was introduced in lesson 3 (Fig. 2.4) using a version of the toilet roll of time adapted to cover fifteen different significant events in the history of life on earth. This activity taught deep time in the way suggested by Catley and Novick (2008) who advocated providing students with knowledge of the correct timing of a small number of critical events together with the visualisation of the relative spacing of these events to provide a framework for greater understanding of evolutionary processes. The resources used in this lesson by the study were developed with permission from existing resources found on the Times Educational Supplement (TES) website (https://www.tes.com) and The Earth Science Teachers Association (https://earthscience.org.uk).

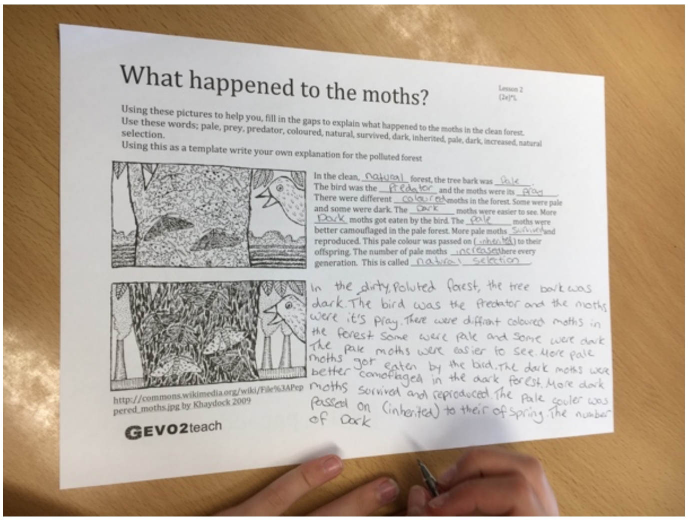
Figure 2.4 Photographs of the main activity from lesson 3; the toilet roll time line, showing that the activity could be carried out either in the classroom or inside along a corridor

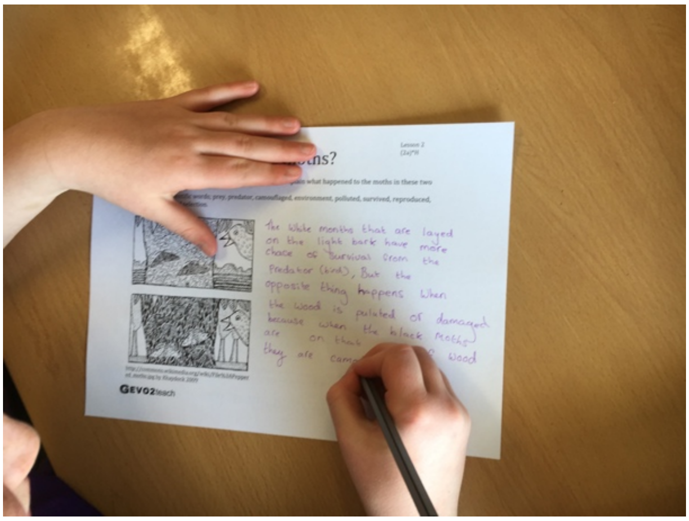
The main activities for lessons 2 and 4 were based on existing age appropriate educational papers. In lesson 2, the Peppered moth (*Biston betularia)* was chosen as the best exemplar species showing natural selection in action (Cook & Saccheri, 2012; Majerus, 2009) Two alternative activities were developed around the predation of Peppered moths by birds to establish whether a hands-on student centred 'hunting' activity was more effective than a seemingly more traditional teacher centred power point activity. The 'hunting' activity (Fig.2.5) was based on the natural selection section of Campos and Sá-Pinto (2013) paper in which they trialled and presented a set of activities that could be used with primary age children. This activity involved the students acting as bird predators using their forceps beaks to 'hunt' the moths on either white or newspaper environments, rather than coloured chocolate prey used in the Portuguese study. Several rounds of timed predation, followed by reproduction of the surviving moths were carried out to show differential survival and increased proportion of mimetic colours. During the activity students were asked to make and test predictions and then to explain the process.

Figure 2.5 A photograph showing a student hunting paper moths in the more student centric hunting investigation which formed one alternative of the main activity in lesson 2.

The alternative moth activity was based on the picture story-book intervention of Kelemen et al. (2014) which was shown to significantly improve the understanding of natural selection in their small scale study of 5 to 8 year-old American children. A PowerPoint presentation explaining the process of natural selection in Peppered moths was developed mirroring the pages taken from a description of their story-book about *pilosas* (a realistic fictional species). After the presentation students were asked to explain the process of natural selection in their own words when provided with diagrams to act as visual stimuli and a glossary of terms. This scaffolded sheet was differentiated so that the students could decide their own appropriate level of difficulty (Fig.2.6). Based on previous literature and some limited small-scale research projects it was expected that the 'hunting' moths activity would be more successful than the scaffolded PowerPoint activity.

Figure 2.6 Photographs showing the nature of the differentiated written exercise carried out after the power point presentation forming the alternative moth main activity in lesson 2. Lower ability students were supplied with keywords and a cloze passage scaffold to explain the first diagram (top photo). Whilst higher ability students were only supplied with the key words with which to explain the diagrams (bottom photo)

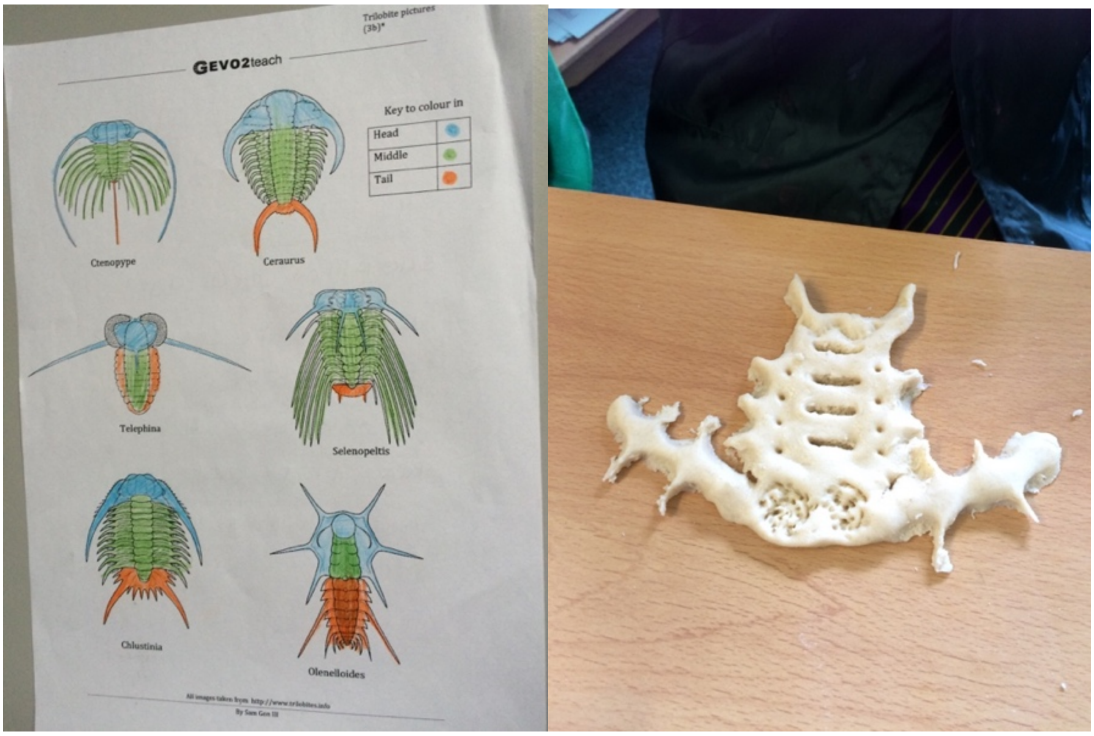




The two alternative activities for lesson 4 both involved the same learning experiences but were developed to establish whether homology and common ancestry were easier to understand if based upon extinct or extant species which also covaried with being related to us as humans or not. Both involved identification of homologous structures and salt dough model making.

The extinct species chosen (Fig 2.7) were various diverse Trilobites adapted from Wagler (2010), whose editorial made suggestions on how to integrate primary age biological and evolutionary content into inquiry based science activities within normal classrooms.

Figure 2.7 Photographs showing the identification of homologous body parts in various Trilobite species and salt dough modelling from lesson 4.



The structure of the pentadactyl limb (Fig 2.8) in a range of tetrapod vertebrate organisms formed the basis of the extant example. This activity was based on the similarities/differences lesson of Nadelson et al. (2009) research which explored the identification of homologous structures by American children between 5 and 7 years old. Using work sheets adapted from the Nuffield Foundation http://www.nuffieldfoundation.org/science-society/activities-evolution, students in this current study identified homologous bones in a range of mammalian examples and then went onto model a human forelimb based on the patterns identified during the lesson.

Figure 2.8 Photographs showing the identification of homologous bones in various extant mammal species and salt dough modelling of a human arm as other alternative main activity in lesson 4.

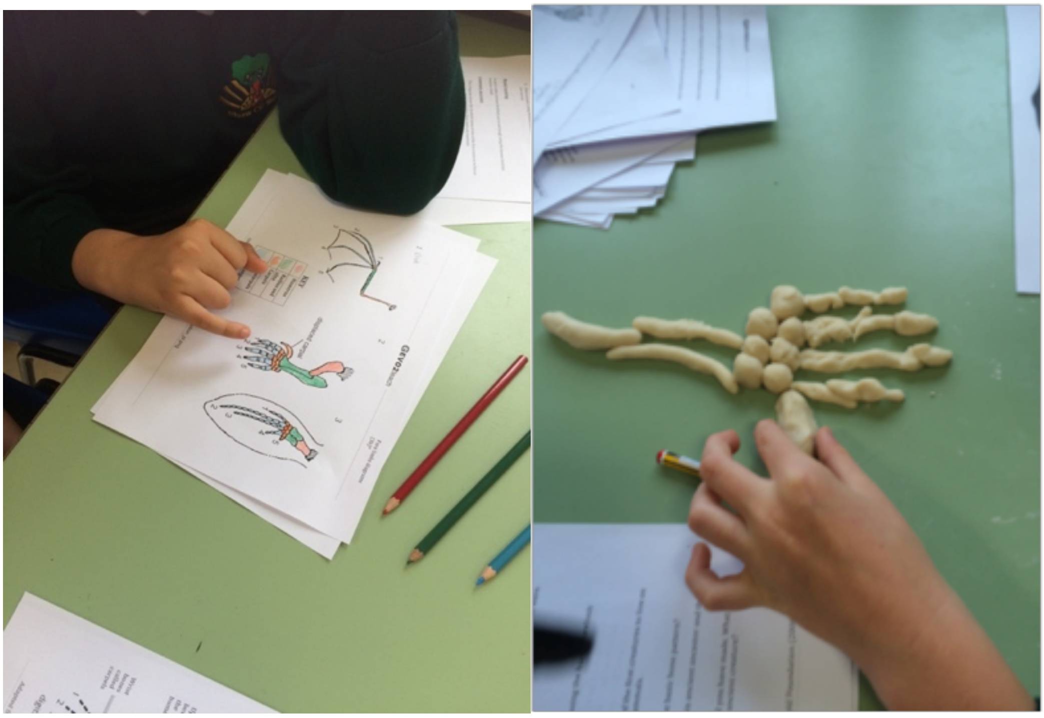


Table 2.2 shows a summary of how the main activities were embedded into the four lessons. For the complete set of all teaching resources please see Appendix F (provided separately due to large file size).

Table 2.2 Summary of the range of different activities into which the main activities were embedded.

Lesson 1	**Variation**
Starter	Power point presentation on variation (definition, examples and causes) and continuum activity to show variation within the class
Main	Quantitative investigation of variation within the class (eye colour, chin dimples and hand span) including mathematical processing of data and transformation into graphs
Plenary	Hospital baby switch scenario exercise to explore understanding of the cause of variation in traits

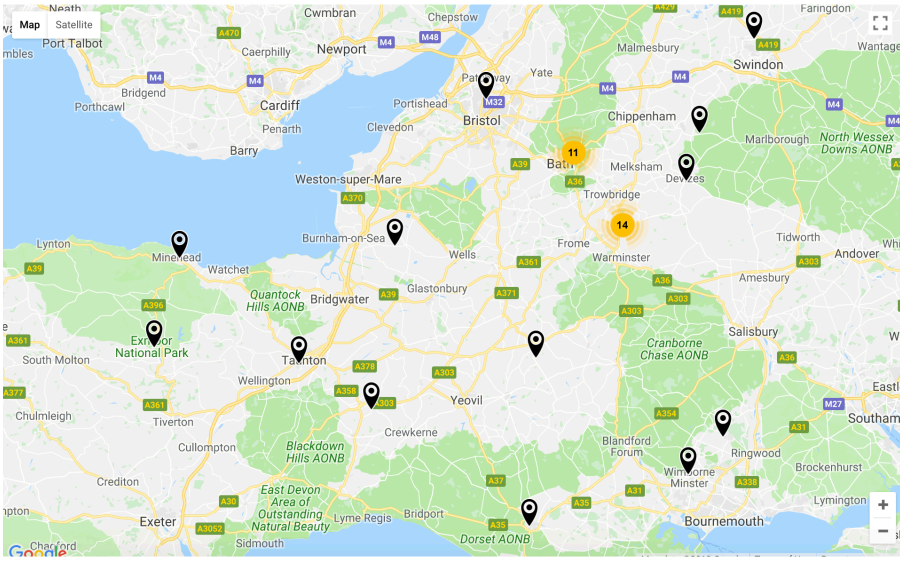
Lesson 2	Natural selection
Starter	Power point presentation on adaptations (polar bear and cactus) and blob game to explore ideas of natural selection, selective advantage, differential survival and extinction
Main	2 alternative activities based on Peppered moth populations as an example of natural selection. Either (a) student centred 'hunting' moths activity carried out in small groups or (b) more traditional teacher centred activity based around a power point presentation and scaffolded written task explaining survival of mimetic peppered moths
Plenary	Story board activity to show process of natural selection in an island dwelling bird population
Lesson 3	Geological time
Starter	Video introducing concept that all organisms are related to a common ancestor using whales as an example
Main	Toilet roll of time activity in which 200 squares of the toilet roll represents 4.6 billion years. Students order and position different cards along the time line to show major geological events from the formation of the earth to the present day emphasizing their relationship to the same common ancestor
Plenary	Mini spiral timeline to hang in class
Lesson 4	Homology and common ancestry
Starter	Power point presentation on how organisms have changed over time and are related to each other
Main	2 alternative activities both based around the identification of homologous structures within related species and model making. Either (a) extant species focusing on the pentadactyl limb in tetrapods or (b) extinct species focusing on *Trilobites*
Plenary	Exercise on phylogenetic trees and the common ancestry of humans

Three homework exercises were also developed to raise awareness of the importance of the contributions made by historical figures (Mary Anning, Jean-Baptise Lamarck and Charles Darwin) to current evolutionary theory. The homework exercises were designed to improve comprehension of the chosen passages as well as numeracy and literacy skills. Completion of the homework was voluntary to match their status in the supplemental information section in the National Curriculum and the fact that not all primary schools set science homework or any homework at all. As this content was optional, care was taken to exclude these topics from any of the assessment items in the student questionnaire.

2.7 School recruitment process was successful

All primary and middle schools within a 50-mile radius of the University of Bath were invited to participate in the study. This distance allowed individual face-to-face contact and teacher training. Please refer to Figure 2.3 showing the distribution of participating schools across the Southwest and the mixture of rural and urban schools recruited.

Figure 2.3 Map showing distribution of the 45 participating schools



The recruitment strategies employed included phone calls, letters, emails and promotional post cards to named teachers within the schools when possible to avoid problems with administration staff acting as ‘gatekeepers’ filtering out contacts for their teaching colleagues. Please see Appendix G for recruitment materials. A success rate of ~10% uptake in the schools contacted was achieved together with a school completion rate of 90%. In Tranche 1 (2016/2017) the sample was collected from 41 separate classes, taught by 37 different teachers in 17 schools. A total of 1152 students completing the pre-test, 988 students completing both pre and post-tests and a much smaller number completing both the post and retention test (n=320). Tranche 2 was slightly larger as collected over two academic years (2016-2018) with a total of 1505 students completing the pre-test, 1309 completing both the pre and post tests and again a smaller number, 523 completing the retention tests. Data in this second tranche was collected from 58 different classes, taught by 48 different teachers in 28 schools (See Table 2.3). All schools in the study were comprehensive schools which did not select students on the basis of ability, only one of the schools in both tranches was an independent school the rest being state schools.

Table 2.3 Summary of participation in the two tranches

	Tranche 1	**Tranche 2**
Number of schools	17	28
Number of classes	40	56
Number of teachers	37	48
Number of students completing pre-test	1152	1505
Number of students completing pre and post-test	988	1309
Number of students completing the post and retention tests	320	523
Number of primary schools	9	24
Number of middle schools	8	4

2.8 Schemes of work were allocated randomly to participating schools

Once a school had agreed to participate one of the four Schemes of Work (SoW) was allocated without any input from the school. This was done carefully in such a way as to form two random trial tests containing a mix of primary and middle schools, locations and approximately equal numbers of students. In schools participating in both tranches the 'opposite' Scheme of Work was allocated so that they delivered the alternative main activities for lessons two and four.

2.9 Participating teachers were given training to use the resources effectively

After random allocation of a Scheme of Work to a specific school, a mutually convenient time slot of least an hour was arranged; lunch time, free time within the school day or after school for the principle researcher to deliver the resources and conduct standardised teacher training. The allocated Scheme of Work was scrutinised lesson by lesson, each individual activity discussed and demonstrated to ensure that all teachers understood how to use the resources and carry out the activities. Guidance was given on how to schedule the lessons to fit into their allocated science lessons, classroom management, the appropriateness of the differentiated tasks and how to overcome potential behavioural/logistical/religious problems. Common alternative conceptions that their students might hold were discussed with reference to the activities provided in the resource packs as well as any questions that arose. This programme of standardised teacher training, detailed teacher information sheets and mark schemes were provided to minimize the possibility of students being taught different things, which would affect measures of student performance.

2.10 Classroom observations were carried out

Classroom observations of the delivery an entire Scheme of Work were also carried out in order to sample the actual teaching and learning taking place in the classroom. The principle researcher observed the teaching of the whole Scheme of Work including assessment in one school per tranche. This enabled observation of the students being taught by their normal teachers using the project's teaching resources in situ and further exploration of timings, practical issues, queries and unexpected problems. See Appendix H for fieldnotes of these lesson observations.

2.11 Feedback on the utility of the allocated resources was collected from the teachers

Qualitative feedback from all participating teachers was obtained when the questionnaires were collected at the end of the topic during a prearranged mutually convenient time slot of at least 1 hour. Individual or small group interviews (when there was more than one teacher in the school) were carried out in all participating schools to allow a detailed investigation from a personal perspective. Data were collected from field notes and around 60 hours semi structured interviews (refer to Appendix J3 for the interview questions asked) which were audio taped and then transcribed verbatim to enhance the understanding of statistical enquiry. Systematic, thematic emergent coding was then carried out in order that this qualitative data could be included in the study to produce a deeper exploration of the research questions, encompassing both objective and subjective standpoints and adding to the richness of the findings.

2.12 Feedback on the questionnaire and resources was collected from student focus groups

Qualitative feedback from a representative sample of students was obtained after the questionnaires had been collected at the end of the topic. The focus group interviews were conducted with small groups of students who had been withdrawn from their classes. The semi structured interviews focused on the student questionnaire; its format, readability, difficulty and whether they understood what the questions meant as well as the resources they remembered. The interviews were audio taped and then transcribed verbatim in order to validate the statistical analysis. (Data not shown)

2.13 Input error checking was carried out

As all of the quantitative data was entered by the principle researcher, analysis of inter-rater reliance was not needed. However, input error checking was carried out by a post doctorate colleague on a representative random sample of 50 student questionnaires (pre, post and retention). When the entered data were compared 18 items out of 900 were found to be incorrectly entered (2%), however, only 0.6% (2/328) of the errors actually affected student score.

2.14 Ethical considerations

Research with children presents special issues as young students are more vulnerable, have fewer legal rights and may not understand the language of informed consent (Fine & Sandstrom, 1988). Appropriate ethical clearance for the project was sought from Departmental Research Ethics Officer by completing an EIRA (Ethical Implications of Research Activity) form in order to start collecting data. Legislation recognises that educational research that involves activities that are within the customary, usual procedures of schools and that involve little or no risk to the participants are exempt from the formal review processes (Fraenkel & Wallen, 2010). Additionally, the subjects were not deceived in any way during the study. For this reason, individual parental permission for the student questionnaire was not necessary as this form of assessment was a normal part of teaching and learning within the classroom. However, each class teacher read out a plain English statement to their students about the project before starting the topic (see Appendix I) outlining their rights to privacy and helping them understand the process enabling their informed consent. Some schools also published this letter to parents to inform them of the study as well.

Prospective participating teachers and their students were informed of the intentions behind the study and were assured that confidentiality and right to privacy would be maintained. They were also assured that there would be no harm to them as individuals and that the results of the study would not influence their grades or performance assessments within school. They were informed that participation in the study was voluntary and that they were free to opt out if they wanted to at any time during the process.

Signed permission was obtained for all audio taped discussions both teacher and student during collection of the qualitative data as this was outside of normal classroom practices. A letter informing the parents and guardians of the students in the focus groups was accompanied by a consent form that parents were requested to complete if they approved of their participation in the focus groups. Only those students whose parents or guardians gave consent were included in this aspect of the study. Additionally, individual permission was obtained from their teachers before their qualitative feedback was collected. (See Appendix J for permission forms). Care was also taken when taking photographs to exclude facial features; where faces appear in this thesis the student's school held pre-existing signed parental permission forms.

2.15 Chapter summary

This chapter has outlined the recruitment of participants and how the teachers were trained to use the resources provided to them by the project. The choice and adaptation of existing assessment items and teaching resources into coherent Schemes of Work has been shown. It has discussed the rationale for mixed methods approach and how different types of data were gathered. The next chapter analyses student level data obtained in both tranches.

Chapter 3 Results of student level analysis comparing both tranches of quantitative data

3.1 Chapter overview

In this chapter the large-scale quantitative data collected in the two tranches of data are analysed side by side. The utility of the assessment instrument is evaluated and validated. Enhanced student performance and retention of knowledge is demonstrated using LOESS regression residuals which correct for pre-test scores and ceiling effect. Possible predictors of student performance; age, gender and relative science ability are examined and the relative effectiveness of the four different Schemes of Work is evaluated.

3.2 Assessment instrument for students was valid and reliable

Before addressing the question of the impact of teaching on student understanding, it is necessary to appraise the utility of the mode of assessment used in this study. A range of different metrics were considered to determine whether the assessment instrument was fit for purpose and the data from both tranches compared to assess its validity and reliability.

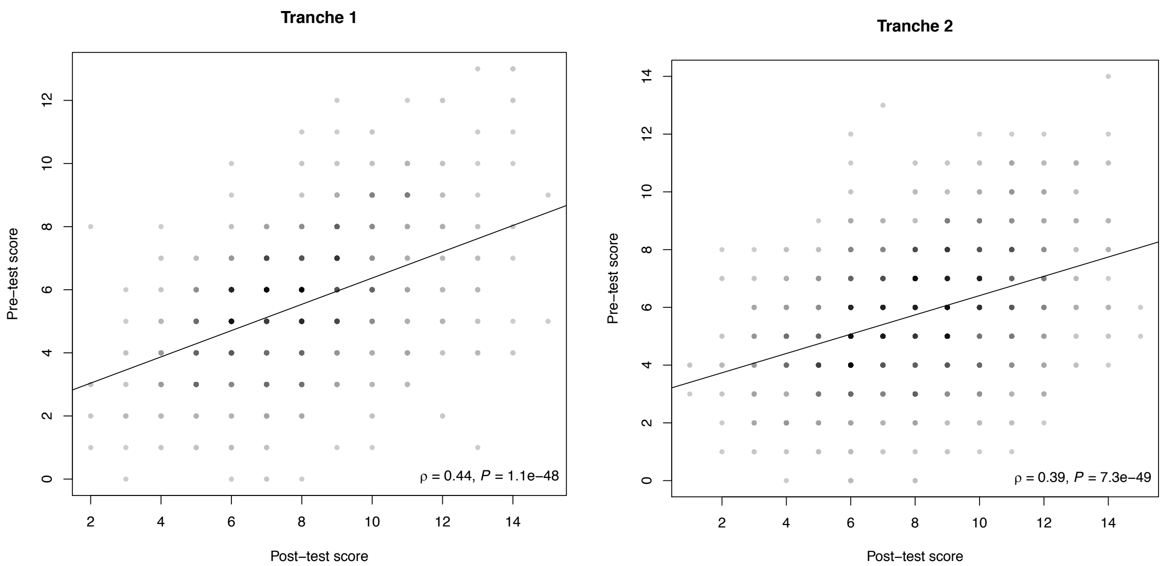
First, to assess internal reliability (or consistency) Cronbach's coefficient alpha was calculated, this value indicating how closely related the set of items are as a group; ranging from 0 (no consistency) to 1 (perfect consistency). Student responses were examined in all three assessment sessions. The pre-tests gave a raw alpha of 'questionable' reliability in tranche 1 (raw $\alpha = 0.66$) but a slightly higher 'acceptable' in tranche 2 (raw $\alpha = 0.73$) suggesting that student responses were slightly less consistent before formal instruction. The coefficients improved in both tranches to an 'acceptable' reliability (raw $\alpha = 0.71$) in tranche 1 and 'good' (raw $\alpha = 0.8$) in the post-tests carried out after instruction. The values for Cronbach's coefficient α for the retention tests remaining 'good' in both tranches and returned very similar values (raw $\alpha = 0.87$ and 0.88). This increase in Cronbach's coefficient showing that the items became more closely related as a group over successive tests, with good repeatability between the two tranches. See Table 3.1 for comparison.

Table 3.1 Comparison of Cronbach's coefficient α values in both tranches of data

Test	Tranche 1	Tranche 2
Pre-test	0.66	0.73
Post-test	0.71	0.80
Retention test	0.87	0.88

Second, under the null hypothesis, that the assessment method was not fit for purpose (e.g. the students are guessing randomly), it would be expected that the post-test and pre-test scores would be uncorrelated. Under the active hypothesis that the mode of assessment is fit for purpose, the paired pre and post-test scores should be correlated for each student. A significant positive correlation between paired student pre and post-teaching scores in both tranches of data ($\rho = 0.44$, $P = < 2.2\text{x}10^{-16}$, Spearman's rank correlation, tranche 1; $\rho = 0.39$, $P = < 2.2\text{x}10^{-16}$, Spearman's rank correlation in tranche 2) was found. With a combined P value of $2.9\text{x}10^{-11}$ (Fisher's test), both tranches showing moderate effect sizes from their correlation coefficients. See Figure 3.1 for comparison of correlation plots. This correlation in both tranches is not obviously compatible with the null hypothesis that the mode of assessment is not fit for purpose and again shows good reliability between data sets

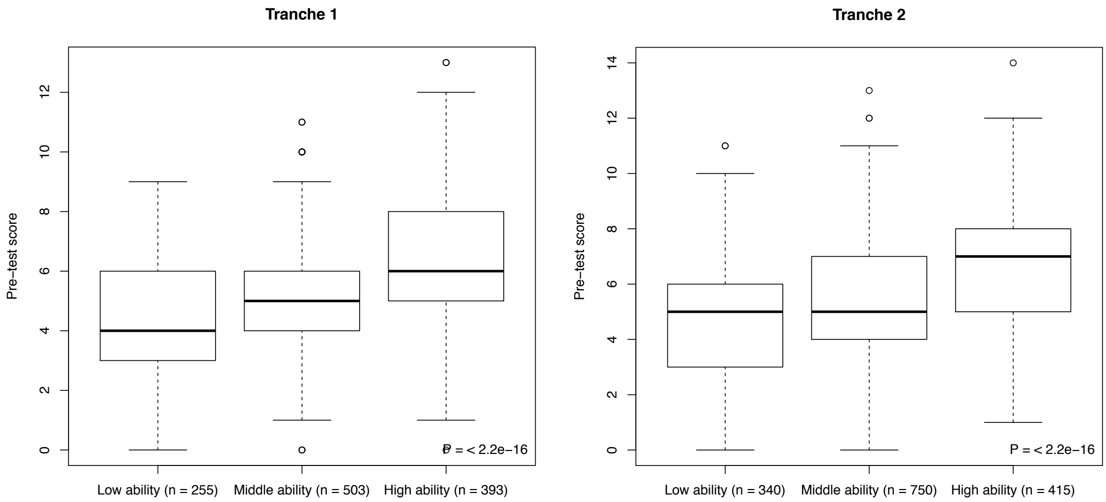
Figure 3.1 Density plots showing correlation between matched pre and post-test scores. The darker the point the more frequently it occurs



Note: Tranche 1, n = 988, mean pre-test score = 5.51 ± 2.15 and post-test score =7.92 ± 2.38. Tranche 2, n = 1309, mean pre-test score = 5.77 ± 2.25 and post-test score = 8.10 ± 2.70.

Third, students identified by their teachers as having higher science ability would be expected to possess greater *a priori* knowledge of the topic and consequently achieve higher marks before formal instruction. A significant positive correlation between teacher assessed science ability and pre-test score in both tranches was found (ρ = 0.34, P = < $2.2x10^{-16}$ Spearman rank correlation, tranche 1; ρ = 0.33, P = < $2.2x10^{-16}$ Spearman rank correlation, tranche 2) with a combined P value of $2.9x10^{-11}$ (Fisher's test), both tranches showing moderate effect sizes. There was also a significant difference in the pre-test scores between all three ability groups in both tranches ($\chi2$ = 136.22, P = 0, Kruskall-Wallis rank sum test, post hoc Dunn test and Bonferroni correction, tranche 1; $\chi2$ = 161.33, P = < $2.2x10^{-16}$, Kruskall-Wallis rank sum test, post hoc Dunn test and Bonferroni correction, tranche 2) both reporting moderate effect sizes (ε^2 = 0.118, in tranche 1; ε^2 = 0.107, tranche 2). The means and medians were also in the expected direction; higher teacher assessed science ability students achieving higher pre-teaching scores (Fig 3.2).

Figure 3.2 Box plots showing the stratification of pre-test score by teacher assessed science ability in both tranches.



Note: Tranche 1, High ability mean = 6.42 ± 2.09, median = 6.00, Middle ability mean = 5.27 ± 1.20, median = 5.00 and Low ability mean = 4.44 ± 1.09, median = 4.00. Tranche 2, High ability mean = 6.84 ± 2.18, median = 7.00, Middle ability mean = 5.55 ± 2.15, median = 5.50 and Low ability mean = 4.73 ± 2.10, median = 5.00.

Fourth, the assessment instrument needs to be accessible and of the correct difficulty. 'Readability' is of great importance as it allows students to access the information within an item and formulate their response. The adapted assessment items were easier to read and more appropriate for this age range compared to the original version, with a mean Flesch reading ease score = 70.04 and Flesch-Kincaid reading grade level = 5.88, compared to 63.21

and 7.17 respectively (Kincaid et al., 1975). The assessment instrument was also of appropriate difficulty for this cohort of students, the mean percentage of correct answers increasing post instruction in both tranches (from 36.56% to 52.83% in tranche 1 and 38.14% to 53.91% in tranche 2) with only a slight decrease in the retention test means (49.18% and 53.52% respectively), these post instruction figures closer to the optimum difficulty score of 62.5% for this type of Multiple Choice Questions (MCQ) (Kaplan & Saccuzzo, 1997). The relative difficulty of each assessment item will be dealt with in the next chapter.

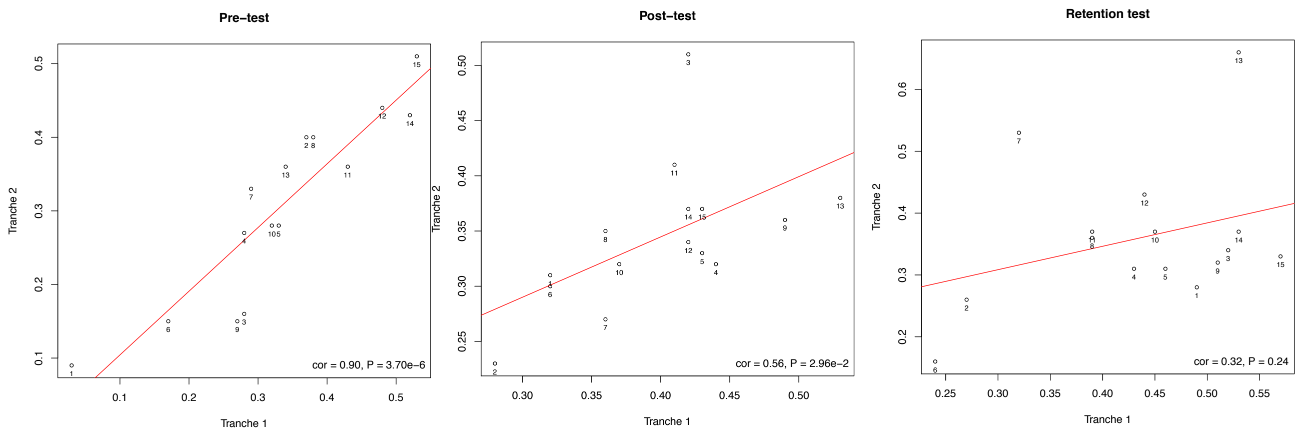
In addition to being of the correct difficulty the assessment instrument must also allow clear discrimination between individual students. Item discrimination, the degree to which students with high overall test scores also got a particular item correct was investigated. Data on individual students and their test scores was used to calculate a total score to separate high and low ordered individuals, upper and lower groups were then defined as the top and bottom 1/3 of the total. The discrimination index was then calculated* and returned for each item ranging between 1.00 to 1.00, with a recommended item discrimination of at least 0.20, larger values indicating that students who got any one item correct also achieved a relatively high score on the overall test. All items returned positive indices with mean values in excess of 0.3 (Table 3.2) in all tests in both tranches suggesting that the assessment instrument was effective at discriminating those students who understood the content from those who do not, rather than just guessing the correct answers. Evidence of repeatability in the results between data sets when the index scores were correlated for each individual item was also found (Figure 3.3).

Table 3.2 Mean item discrimination index scores for all tests in both tranches

Tranche	Pre-test	Post-test	Retention test
1	0.31 ± 0.12	0.34 ± 0.06	0.36 ± 0.12
2	0.33 ± 0.13	0.40 ± 0.07	0.44 ± 0.10

*Note: *calculated using the formula [number correct in the upper group - number correct in the lower group) / size of each group].*

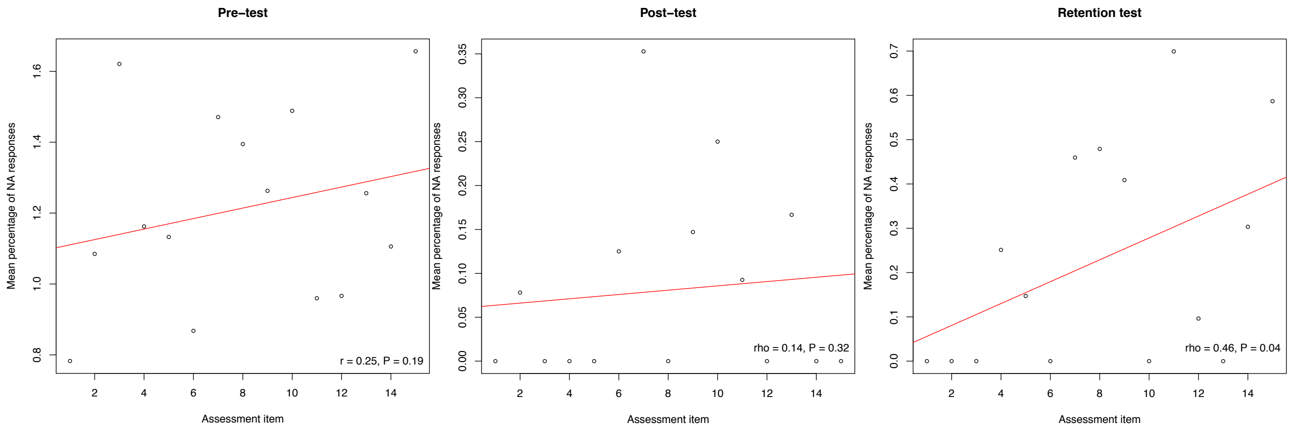
Figure 3.3 Scatterplots with lines of regression showing correlation between item discrimination index values in all three assessment sessions.



Notes: There were significant positive correlations between index values in the two tranches in the pre-test (r = 0.90, P = 3.70x10^{-6}*, Pearson's correlation) and the post-test (r = 0.56, P = 2.96x10*$^{-2}$*, Pearson's correlation). A positive correlation with a medium effect size (r = 0.32, P = 0.24, Pearson's correlation) was also obtained in the retention test although this correlation not significant.*

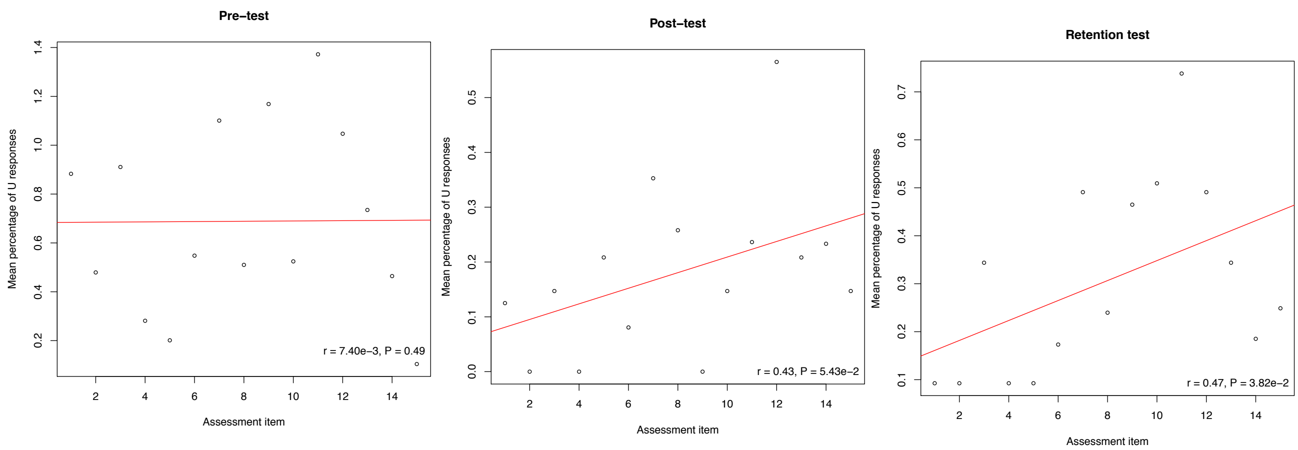
Finally, student completion of the assessment items was considered as a function of item number to check for any evidence of question fatigue which could negatively impact upon the results of items appearing later in the assessment instrument. The correlation between mean class percentage of null (NA) and ambiguous (U) responses and assessment item number was conducted using the combined data from both tranches of students who completed all three assessments (pre, post and retention), to allow direct comparisons to be drawn. One tailed statistical tests were employed as a positive correlation between assessment item number and the percentage of NA and U responses would be expected if question fatigue occurred within the test sessions. There were no significant correlations between assessment item number and the mean class percentage of null (NA) or ambiguous (U) responses in either the pre or post assessment sessions, all with small effect sizes except the post result for U responses which had a moderate effect size (Figures 3.4, 3.5). However, a significant correlation between assessment item number and the percentage of both NA and U responses was found, with moderate effect sizes in both retention tests. These results suggest that although there were positive correlations between increasing assessment item number and the mean class percentage of NA and U responses in all assessment sessions the effect of question fatigue was not significant confounding factor in the of the pre and post assessment session results on which the majority of our analysis was focused. However, the effect of question fatigue negatively impacting on the results of the retention test cannot be ruled out and could explain some of the waning effect encountered within individual assessment items, although the percentage of NA/U responses were extremely low.

Figure 3.4 Scatterplots with lines of regression showing correlation between mean class percentage of null (NA) responses and assessment item number in all three assessment sessions.



Notes: Mean class percentage (n = 40) of NA responses calculated from data obtained from only students completing all three assessments in 40 classes. Pre-test mean % = 1.21 ± 0.29, median = 1.16, r = 0.25, P = 0.19, Pearson's correlation coefficient. Post-test *mean % = 0.09* ± 0.11, median = 0.00, ρ = 0.14, P = 0.32, Spearman's rank correlation coefficient. Retention test *mean % = 0.23* ± 0.25, median = 1.15, ρ = 0.46, P = 0.04 Spearman's rank correlation coefficient.

Figure 3.5 Scatterplots with lines of regression showing correlation between mean class percentage of ambiguous or undecided (U) responses and assessment item number in all three assessment sessions.

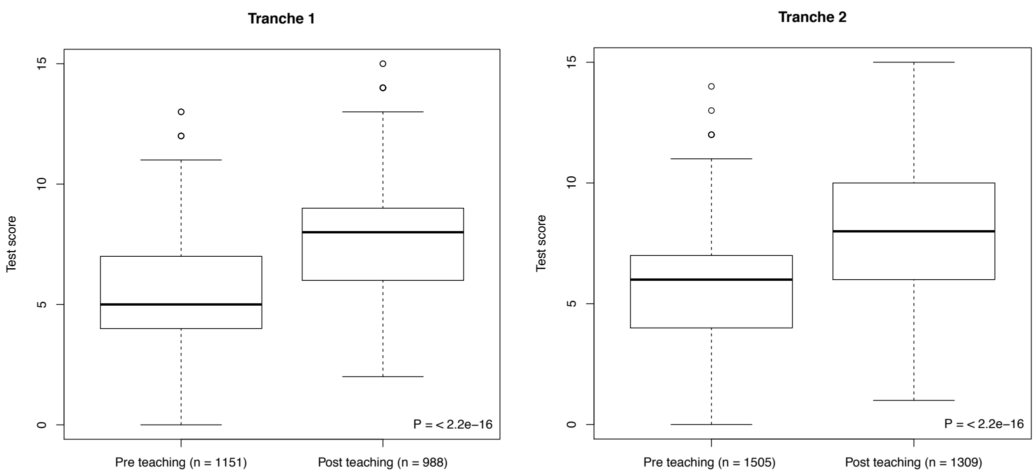


Notes: Mean class percentage (n = 40) of U responses calculated from data obtained from only students completing all three assessments in 40 classes. Pre-test mean % = 0.69 ± 0.38, median = 1.16, r = 7.4x10^{-3}, P = 0.49, Pearson's correlation coefficient. Post-test mean % = 0.18 ± 0.15, median = 1.15, r = 0.43, P = 5.43x10^{-2}, Pearson's correlation coefficient. Retention test mean % = 0.31 ± 0.20, median = 0.25, r = 0.47, P = 3.82x10^{-2}, Pearson's correlation coefficient.

3.3 Teaching interventions significantly improved student performance using multiple measures

Given that the mode of assessment is both reliable and valid, exploration of whether the teaching intervention packages provided by the project can improve student understanding is warranted. This issue can be addressed in two modes. First, all pre-test and post-test scores can be compared in an unpaired manner. This analysis found that the mean student test score increased significantly in tranche 1 by 2.44 marks (16.27%) between the pre and post-test, (w = 254800, $P < 2.2\times10^{-16}$, Wilcoxon rank sum test) with a large effect size (Cliff's d = 0.55) as well as in Tranche 2, where there was an increase 2.37 marks (15.80%) between the pre and post-test (w = 505360, $P < 2.2\times10^{-16}$, Wilcoxon rank sum test), again with a large effect size (Cliff's d = 0.49); Fig. 3.6.

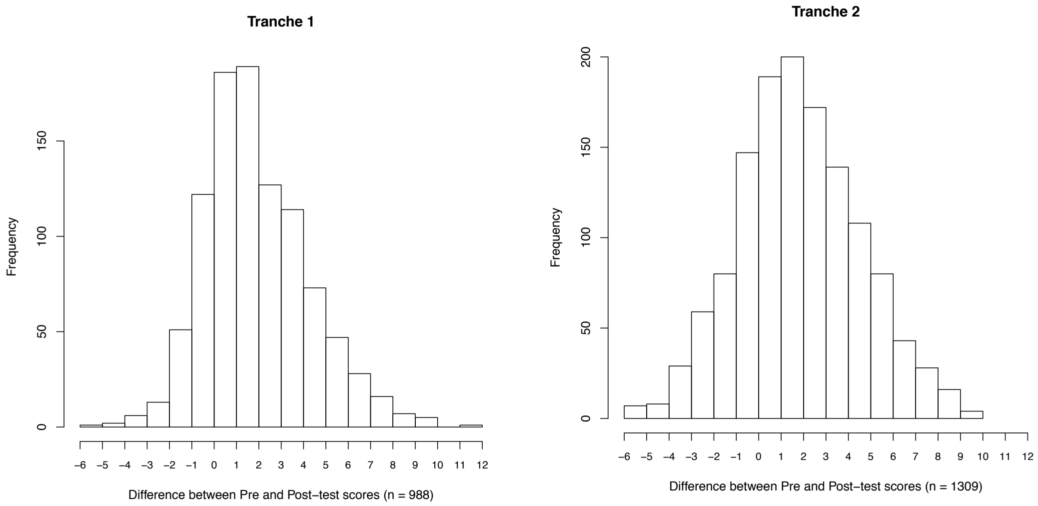
Figure 3.6 Box plots showing the difference between unpaired pre and post-tests in both tranches of data for all students who participated in the study.



Notes: pre-test mean = 5.48 ± 2.13, post-test mean = 7.92 ± 2.38, tranche 1; pre-test mean = 5.72 ± 2.27, post-test mean = 8.09 ± 2.70, tranche 2.

As this analysis doesn't control for the performance of any given student, the second mode of analysis considers the distribution of the change in score values for all students who took both the pre and post-test assessments; Fig. 3.7. When pre and post-test scores were analysed in a paired manner the mean student score increased significantly by 2.42 marks (16.13%) between the pre and post-test, (v = 15552, $P = < 2.2\times10^{-16}$, Wilcoxon signed rank test) with a large effect size (Cliff's d = 0.55) in tranche 1. Compared to a significant increase of 2.32 marks (15.47%) between pre and post-test, (v = 658326, $P = < 2.2\times10^{-16}$, Wilcoxon signed rank test) again with a large effect size (Cliff's d = 0.48) in tranche 2.

Figure 3.7 Histograms showing the distribution of the change in score for all students taking both the pre and post-tests in both tranches.

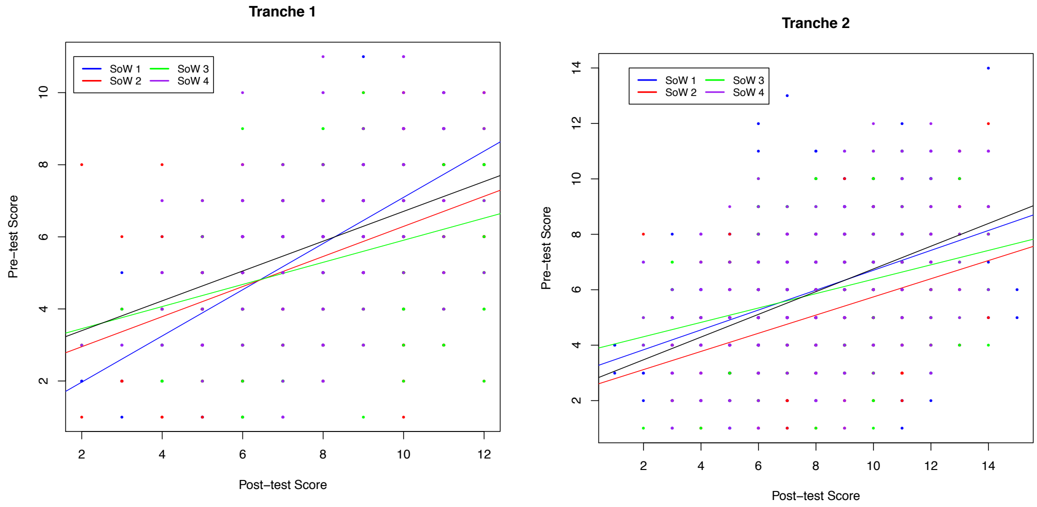


Notes: pre-test mean = 5.51 ± 2.15, post-test mean = 7.92 ± 2.38, tranche 1; pre-test mean = 5.77 ± 2.25, post-test mean = 8.09 ± 2.70, tranche 2.

In an educational setting, effect size is another way to assess the effectiveness of a particular intervention as it enables the measurement of both the *improvement* (gain) in learner achievement within a group of learners and the *variation* of student performances expressed on a standardised scale. From his analysis of hundreds of international and national educational interventions, Hattie (2012) determined a 'hinge point' effect size of 0.4 could be used as a guide to the effectiveness of educational interventions. Effect sizes above 0.4 being effective and those lower than 0.4 needing further consideration or modification. The effect sizes for both tranches of data (0.55 in tranche 1 and 0.48 in tranche 2) were above the 0.4 'hinge point' indicating that the teaching intervention programmes were effective and worth using.

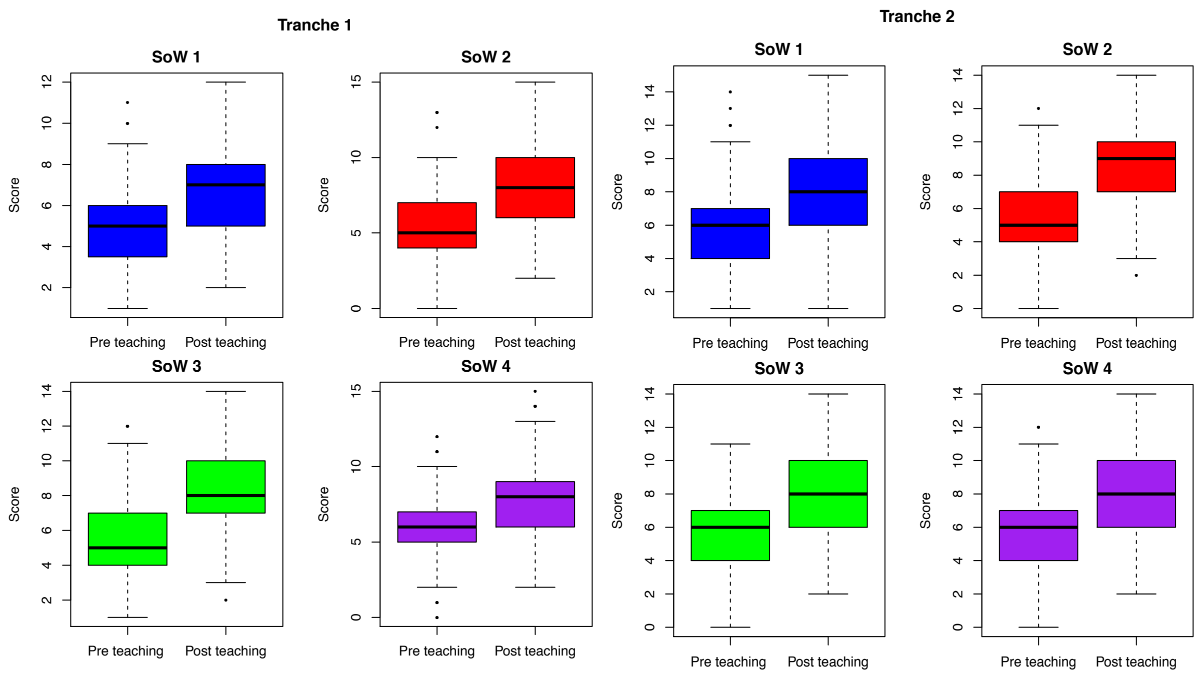
The impact of each of the four Schemes of Work on student performance was also considered using matched pre and post-test scores. When pre and post-test scores were compared, significant positive correlations ($\rho > 0.3$, $P < 0.0001$, Spearman's rank correlation) with moderate effect sizes were obtained for each of the four separate Schemes of Work in both tranches of data (Fig. 3.8) together with a significant difference ($P < 2.2 \times 10^{-16}$, Wilcoxon signed rank test) between the pre and post-test scores for all Schemes of Work in both tranches; Fig.3.9, with moderate to large effect sizes (ε^2 values ranging from 0.11 to 0.46). These results suggest that the instruction provided by all four different Schemes of Work significantly improved student performance.

Figure 3.8 Coloured density plots with lines of regression showing correlation between matched pre and post-test scores stratified by Scheme of Work (SoW) in both tranches.



Notes: Tranche 1; Scheme of Work 1, n = 171, pre-test mean = 5.89 ± 2.30, post-test mean = 7.74 ± 2.78. Scheme of Work 2, n = 314, pre-test mean = 5.25 ± 2.17, post-test mean = 8.49 ± 2.54. Scheme of Work 3, n = 241, pre-test mean = 5.95 ± 2.15, post-test mean = 8.36 ± 2.60. Scheme of Work 4, n = 262, pre-test mean = 5.91 ± 2.29, post-test mean = 7.95 ± 2.76. Tranche 2; Scheme of Work 1, n = 422, pre-test mean = 5.11 ± 2.51, post-test mean = 6.91 ± 2.13. Scheme of Work 2, n = 288, pre-test mean = 5.49 ± 2.25, post-test mean = 8.09 ± 2.57. Scheme of Work 3, n =278, pre-test mean = 5.39 ± 2.06, post-test mean = 8.33 ± 2.34. Scheme of Work 4, n = 321, pre-test mean = 5.89 ± 2.10, post-test mean = 8.02 ± 2.17.

Figure 3.9 Box plots showing the difference between matched pre and post-tests in both tranches of data stratified by Scheme of Work (SoW)

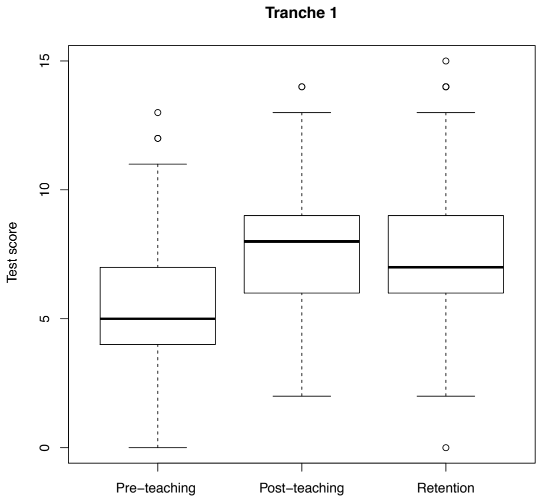


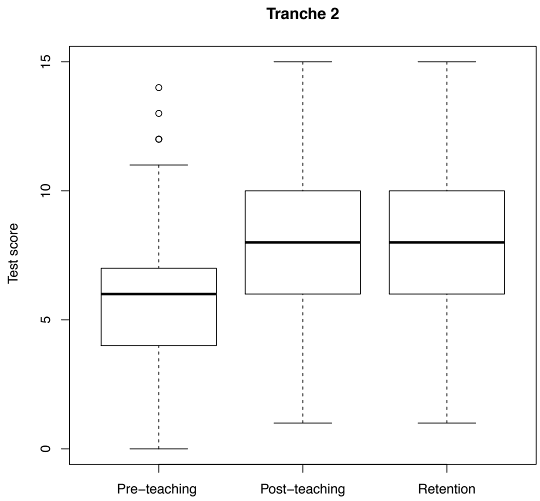
3.4 Evidence of longer-term retention and some waning

A common feature of many teaching interventions is that they lead to short term improvements in understanding, but such improvements are subsequently lost. To address the issue of longer-term retention, the results from a more limited sample obtained from students who took all three tests was considered ($n = 320$ in tranche 1 and $n = 523$ in tranche 2). The period between the post and the retention tests were similar for both tranche (mean 130.93 ± 73.20 days, tranche 1; 134.96 ± 64.78 days, tranche 2). The data was analysed to evaluate whether there was evidence of learning retention or waning.

If there is some degree of longer-term retention, it would be expected that the retention scores should be significantly higher than the pre-test scores. A highly significant difference between pre and retention test scores was found ($P = 2.1 \times 10^{-14}$, Kruskal Wallis rank sum test, post hoc Nemenyi's test, tranche 1; $P < 2.2 \times 10^{-16}$ in tranche 2, Kruskal Wallis rank sum test, Nemenyi's post hoc, tranche 2), with a combined P value of 2.86×10^{-11} (Fisher's test). Second, if there is a waning effect, whereby over time gains made are gradually lost, the post teaching scores would be expected to be higher than retention scores. This was found to be the case in tranche 1 ($P = 0.0036$, Kruskal Wallis rank sum test, post hoc Nemenyi's test) but not so in tranche 2, where there was no significant difference between the post and retention scores ($P = 0.50$, Kruskal Wallis rank sum test, post hoc with Nemenyi's test; Fig. 3.10). However, these results give a statistically significant combined P value ($P = 1.31 \times 10^{-2}$ Fisher's test) both tranches showing moderate effect sizes ($\varepsilon^2 = 0.17$, tranche 1, $\varepsilon^2 = 0.13$, tranche 2) suggesting that the teaching interventions had some degree of long-term retention but also that some of this understanding waned over time.

Figure 3.10 Box plots comparing matched pre, post-teaching and retention scores in both tranches of data.





Note: Tranche 1, n = 320, pre-test mean = 5.50 ± 2.16, median = 5.00, post-test mean = 7.90 ± 2.26, median = 8.00, retention test mean = 7.34 ± 2.52, median = 7.00. Tranche 2, n = 523, pre-test mean = 5.91 ± 2.24, median = 6.00, post-test mean = 8.19 ± 2.86, median = 8.00, retention test mean = 8.03 ± 2.92, median = 8.00.

3.5 LOESS residuals were used to adjust pre-test scores

Before any analysis involving the difference between pre and post-test scores was carried out the data were adjusted by correcting for pre-test scores. Unlike measures of normalised gain which make assumptions about the relationship between the change in score and the pre-test score, this study employed a method that derives the relationship from a regression curve. The residuals of a LOESS curve (a locally weighted polynomial regression for non-parametric data) for the change in student score against pre-test score was used for all future data analysis. This method also mitigates some of the 'ceiling effect' problem present in any form of quantitative assessment, in which students with higher pre-test scores can only make smaller relative gains compared with lower scoring students as their initial score is already high, there being a maximum mark or ceiling to the test. Likewise, students with lower initial scores have the potential to make larger relative gains but also smaller relative losses. Analysis of the uncorrected change in score with pre-test score confirming that this ceiling effect is pertinent to both data sets as we found a significant negative correlation of moderate effect size in both tranches ($\rho = -0.438$, $P = < 2.2 \times 10^{-16}$, Spearman's rank correlation, tranche 1; $\rho = -0.403$, $P = < 2.2 \times 10^{-16}$, Spearman's rank correlation, tranche 2).

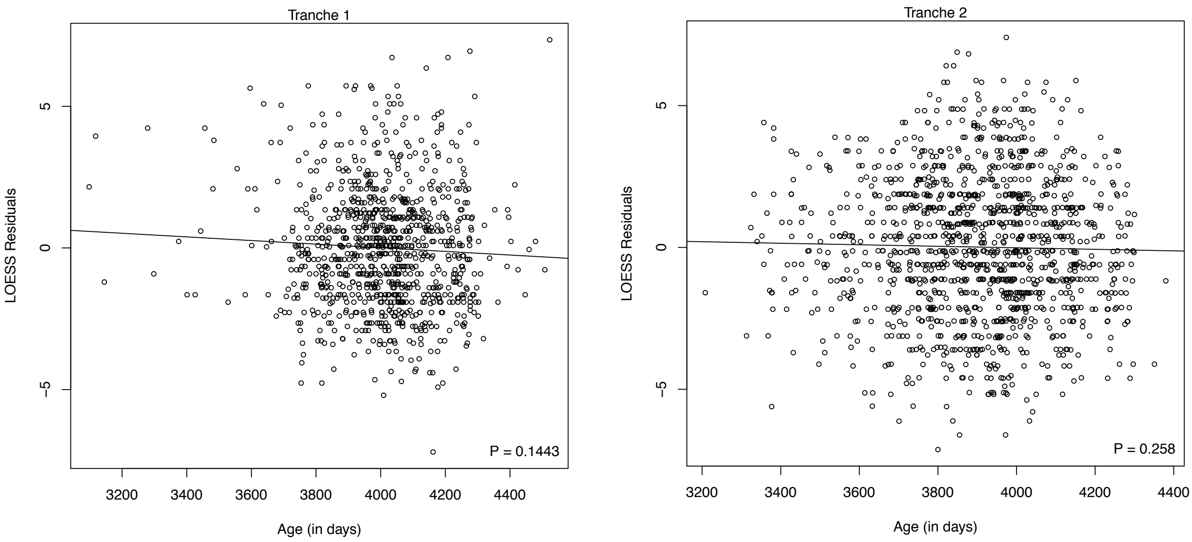
Additionally, this method also corrected for the uneven distribution of raw pre-test scores matched for post-test scores when stratified by Scheme of Work. This distribution of raw data was found to be significantly different across the four different Schemes of Work in both tranches ($\chi 2$ = 16.27, P = 9.9×10^{-4}, n = 988, Kruskal Wallis rank sum test, with Bonferroni correction, tranche 1; $\chi 2$ = 20.59, P = 1.3×10^{-4}, n = 1309, Kruskal Wallis rank sum test, with Bonferroni correction, tranche 2). With a combined P value of P =2.1×10^{-6} (Fisher's test) and small effect sizes ($\varepsilon^2 = 0.017$, tranche 1, $\varepsilon^2 = 0.016$, tranche 2).

3.6 Only teacher assessed science ability consistently predicts student response to teaching evolution

If the distribution of the adjusted change in score values are considered, there is still considerable variation to be explained at the individual student level. Which factors might predict variation in performance? Could gender predict response to teaching? Some evidence suggests that primary school students think of science as a masculine subject (Archer, 1992). According to this stereotype boys are better at science than girls which is often the implicit message hyped in media headlines (Barnett & Rivers, 2004), leading to girls having a less positive attitude towards science and a reduced perception of their own ability in the subject (Heller & Ziegler, 1996). However, other evidence from the results of meta-analyses (Greany et al., 2016) report comparable gender performance in UK primary school students supporting the gender similarities hypothesis (Hyde, 2005) in which boys and girls are similar in most psychological variables. Or will the data confirm that the trend for girls out performing boys at GCSE (Bramley et al., 2015) also occurs in primary schools? Likewise, it might be expected that students identified by their teachers as having higher ability in science relative to their peers would achieve a larger change in marks after instruction. Finally, it has been suggested that older students may be cognitively more able to grasp this difficult topic especially the more abstract concepts such as homology and deep time. Thus, this study considers these three parameters utilising the adjusted change in scores from the LOESS residual scores.

No evidence that student age has any explanatory power was found, as there was no significant difference in the adjusted change in score values (LOESS residual scores) and student age (taken on the date of the pre-test) in either tranche of data (ρ = -0.046, P = 0.144, Spearman's rank correlation, tranche 1; ρ = -0.031, P = 0.258, Spearman's rank correlation, tranche 2; Fig. 3.11). Both tranches reporting small effect sizes from their correlation coefficients with a combined P value of 0.16 (Fisher's test).

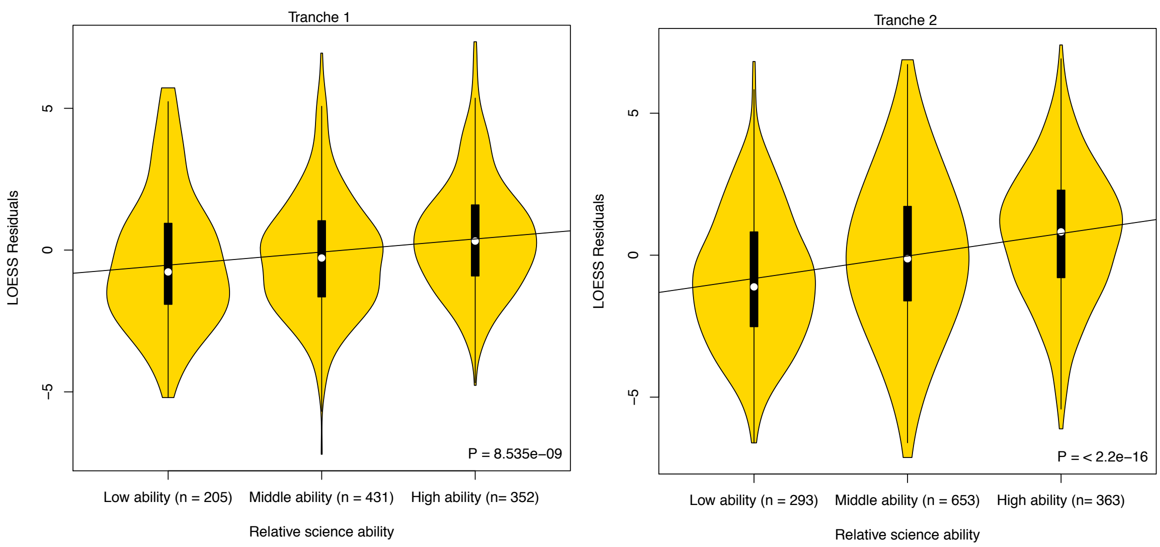
Figure 3.11 Scatterplots with lines of regression showing correlation between adjusted LOESS residual scores and student age taken on the day of the pre-test.



Note: Tranche 1, n = 988, mean age = 4009 days ± 173.97, median = 4017; Tranche 2, n = 1309, mean age = 3902 days ± 188.64, median = 3917.

There was, by contrast, a significant difference with small effect sizes between high, middle and low ability levels ($\rho = 0.182$, $P = 8.5 \times 10^{-9}$, Spearman's rank correlation, tranche 1; $\rho = 0.226$, $P = < 2.2 \times 10^{-16}$, Spearman's rank correlation, tranche 2; Fig. 3.12), with higher ability students showing a greater adjusted change in score.

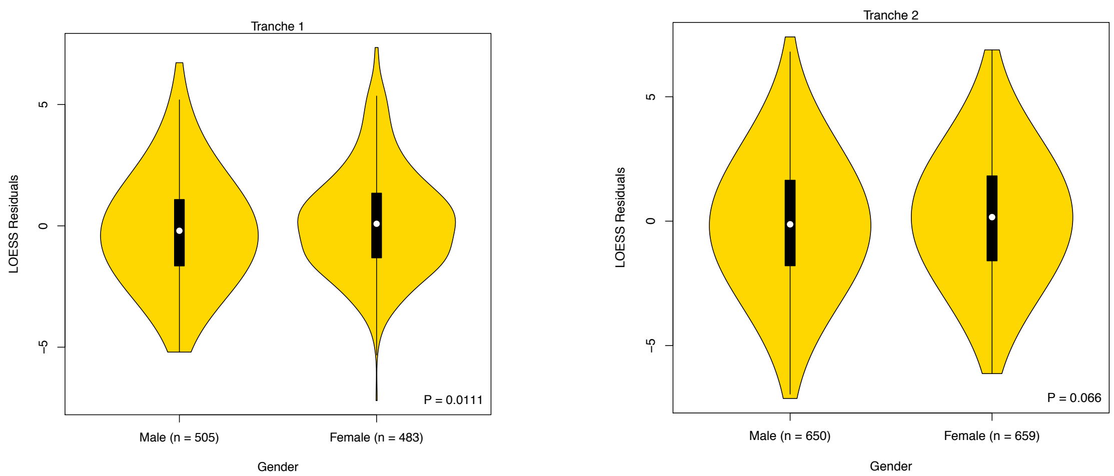
Figure 3.12 Violin plots with lines of regression showing adjusted LOESS residual scores stratified by relative student ability in science.



Note: Tranche 1, mean low ability = -0.38 ± 2.34, median = -0.77, mean middle ability = -0.22 ± 2.03, median = -0.28, mean low ability = 0.48 ± 1.95, median = 0.32; Trance 2, mean low ability = -0.87 ± 2.32, median = -1.12, mean middle ability = 0.01 ± 2.48, median = -0.13, mean low ability = 0.73 ± 2.34, median 0.82.

When the adjusted LOESS residual scores were stratified for gender, a significant result (w = 1333, P = 0.01, Wilcoxon test) was obtained, with female students achieving higher adjusted change in score values in tranche 1. However, there was no significant difference between gender specific performance in tranche 2 (w = 2267, P = 0.07, Wilcoxon test). These values gave a significant combined P value of 6.00×10^{-3} (Fisher's test) remaining significant after Bonferroni correction for four different tests (i.e. P < 0.0125), however, both had negligible effect sizes (Cliff's d = 0.09, tranche 1; Cliff's d = 0.06, tranche 2). This suggests whist gender appears to be a statistically significant predictor of student performance, the effect sizes reveal that the effect of gender is negligible for our relatively larger sample sizes. See Figure 3.13.

Figure 3.13 Violin plots showing adjusted LOESS residual scores stratified by gender



Note: Tranche 1, mean male = -0.18 ± 2.04, median -0.20, mean female = 0.18 ± 2.15, median = 0.09; Tranche 2, mean male = -0.12 ± 2.48, median = -0.13, mean female = 0.14 ± 2.45, median = 0.16.

A multivariate regression model performed on the combined effect of gender, age and ability was also found to be significant and accounting for 3.30% of the total variance in student score across both tranches of data (adjusted $R^2 = 3.30 \times 10^{-2}$, $P = 6.72 \times 10^{-8}$, tranche 1; adjusted $R^2 = 5.30 \times 10^{-2}$, $P = 4.72 \times 10^{-16}$, tranche 2). The results confirming that age was not significant (coefficient = -1.00×10^{-3}, P = 0.07, tranche1; coefficient = -5.00×10^{-4}, P = 0.13, tranche 2), adjusted scores increased significantly with increasing science ability (coefficient = 0.46, P =

3.00×10^{-7}, tranche1; coefficient = 0.80, P = < 2.2×10^{-16} , tranche 2), and that girls scored significantly higher than boys in tranche 1 (coefficient = 0.34, P = 0.01) but not significantly higher in tranche 2 (coefficient = 0.24, P = 0.08).

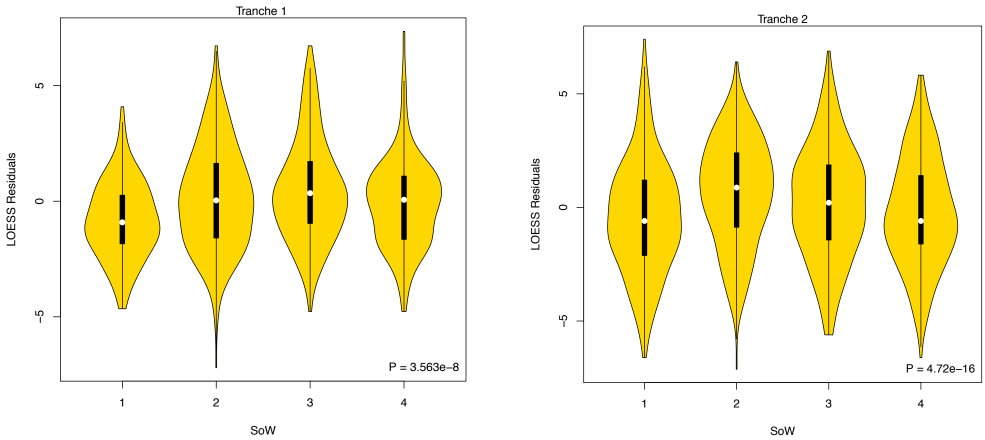
3.7 Some Schemes of Work were more effective than others

One of the main aims of this study was to investigate the most effective ways of increasing understanding of evolutionary concepts among 9-11 year olds. To this end the adjusted LOESS residual scores were stratified by Scheme of Work and then a Kruskal-Wallis rank sum test with a Bonferroni correction followed by a Dunn post hoc test was carried out, in order to compare the relative effectiveness of the four different Schemes of Work (see Fig.3.14). This analysis gave statistically significant results in both tranches ($\chi 2 = 37.53$, P = 3.56×10^{-8}, tranche 1; $\chi 2 = 40.91$, P = 6.84×10^{-9}, tranche 2). In tranche 1, the order of increasing relative effectiveness of the Schemes of Work was; 1, 4, 2 with 3 being the most effective. Scheme of Work 1 was found to be significantly less effective than all of the other Schemes of Work, whilst the other three schemes were equally as effective there being no significant difference between them. Similarly, in tranche 2, Scheme of Work 1 was also the least effective followed by in order of increasing effectiveness; 4, 3 with Scheme of Work 2 being the most effective. However, the significance of these results is not as straight forward as in tranche 1. There was no significant difference between Schemes of 2 and 3 (P = 0.06), 3 and 4 (P = 0.27) or between 1 and 4 (P = 1.00). However, there was a significant difference between Schemes of Work 1 and 2 (P = 8.53×10^{-9}), 1 and 3 (P = 1.88×10^{-3}) and 2 and 4 (P = 1.84×10^{-5}). Both sets of results returning small effect sizes ($\varepsilon^2 = 3.80 \times 10^{-2}$, tranche 1; $\varepsilon^2 = 3.13 \times 10^{-2}$ tranche 2). Overall, the results of these statistical analyses suggest that Scheme of Work 1 was found to be the least effective instruction package relative to the other three schemes, Scheme of Work 4 was the next least effective whilst the other two schemes seemed to have fairly similar relative effectiveness (see Table 3.3 for a summary). However, it should be noted that even though Scheme of Work 1 was the least effective relative to the other schemes it also significantly improved student performance (see section 3.3).

Table 3.3 summary of the order of effectiveness for the 4 Schemes of Work in both tranches

Order of effectiveness	Tranche 1	Tranche 2
Most ↓	**3**	**2**
	2	**3**
	4	**4**
Least	**1**	**1**

Figure 3.14 Violin plots showing adjusted LOESS residuals stratified by Scheme of Work (SoW) in both tranches.



Note: Tranche 1, Scheme of Work 1 mean = -0.83 ± 1.65, median = -0.91; Scheme of Work 2 mean = 0.16 ± 2.25, median = 0.04; Scheme of Work 3 mean = 0.47 ± 2.19, median = 0.35; Scheme of Work 4, mean = -0.07 ± 1.93, median = 0.06. tranche 2, Scheme of Work 1 mean = -0.40 ± 2.51, median = -0.60; Scheme of Work2 mean = 0.67 ± 2.34, median = 0.85; Scheme of Work 3 mean = 0.198 ± 2.48, median = 0.21; Scheme of Work 4 mean = -0.191 ± 2.40, median = -0.60.

If a similar multivariate model on the combined effect of gender, age, ability and Scheme of Work is carried out to confirm the findings and to ascertain whether any more of the variance can be explained by including the Scheme of Work as a factor, slightly more (7.10% compared to 3.30%) of the total variance in student score across both tranches of data can be accounted for (adjusted R^2 = 7.1x10^{-2}, P = 8.67x10^{-15}, tranche 1; adjusted R^2 = 8.30x10^{-2}, P = < 2.2x10^{-16}, tranche 2). These results also confirm that age was not significant (coefficient = -6.61x10^{-5}, P = 0.86, tranche1; coefficient = -1.68x10^{-5}, P = 0.96, tranche 2), adjusted scores increased significantly with increasing teacher assessment of student science ability (coefficient = 0.48, P = 7.93x10^{-8}, tranche1; coefficient = 0.83, P = < 2.2x10^{-16} , tranche 2), and that girls scored significantly higher than boys in tranche 1 (coefficient = 0.35, P = 7.00x10^{-3}) but not significantly higher in tranche 2 (coefficient = 0.24, P = 6.40x10^{-2}). This analysis also confirms that Scheme of Work 1 was significantly less effective relative to the other schemes in tranche 1 (coefficient = 0.97, P = 9.24x10^{-7} Scheme of Work 2; coefficient = 1.32, P = 3.48x10^{-10} Scheme of Work 3; coefficient = 0.69, P = 0.001 Scheme of Work 4) after Bonferroni correction. Scheme of Work 1 was also the least effective relative to the other schemes in tranche 2, however, it was only significantly less effective than scheme 2 (coefficient = 1.15, P = 3.61x10^{-10}) and scheme 3 (coefficient = 0.64, P = 5.00x10^{-4}) but not scheme 4 as there was no significant difference between them (coefficient = 0.13, P = 0.46).

These results confirm and strengthen the previous findings and suggest that Schemes of Work 2 and 3 are the most effective teaching intervention programmes.

3.8 Pairs of lesson activities interact in a reciprocal and reinforcing manner

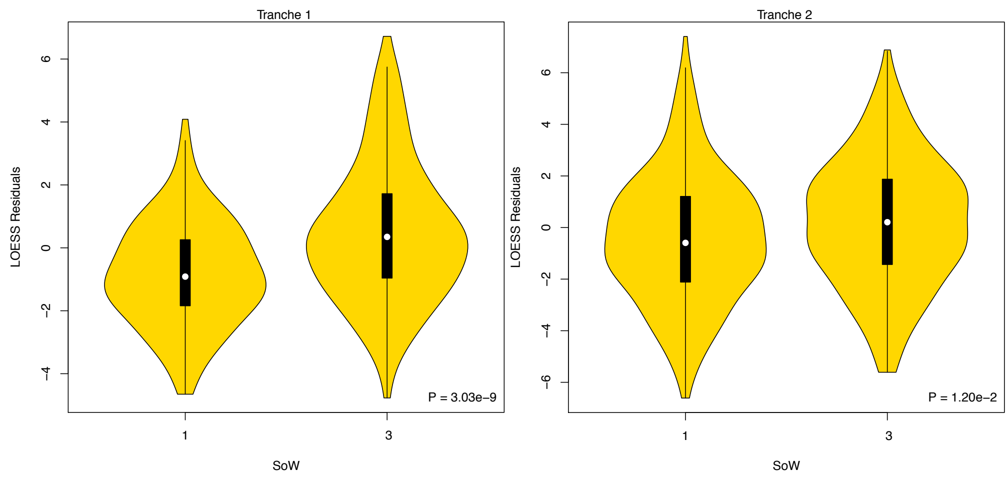
The arrangement of the lesson 'mains' contained within lessons 2 and 4 of the study was carefully designed to allow the determination of the most effective single activities, see Table 4.3. By comparing two different pairs of Schemes of Work (SoW); 1 and 3 and 2 and 4 the relative effectiveness of the different peppered moth activities in lesson 2 can be ascertained, followed by the different homology/common ancestry tasks in lesson 4. Scheme of Work 3 was found to be significantly more effective than Scheme of Work 1 in both tranches (see Fig. 3.15) with a combined P = 2.31×10^{-7} (Fisher's test), suggesting that the moth power point activity was more effective than hunting paper moths if followed by studying common ancestry and homology in Trilobites. However, when Schemes of Work 2 and 4 were compared to assess which moth activity was more effective when followed by homology and common ancestry using the pentadactyl limb it was found that the hunting moths activity was more effective in both tranches but only significantly so in tranche 2 (see Fig. 3.16) with a combined P = 1.15×10^{-5} (Fisher's test). These results suggest that there were positive and significant interactions between the different moth activities in lesson 2 and the homology/common ancestry tasks in lesson 4. Completing the power point moth activity followed by studying Trilobites or conversely completing the hunting moths activity followed by studying the pentadactyl limb were most effective in improving student performance, with the other combinations being less effective.

Table 3.3 Comparison of the main activities carried out by the students in lessons 2 and 4.

Scheme of Work	Main activity carried out by students	
	Lesson 2	Lesson 4
1	Hunting paper moths	Trilobites
2	Hunting paper moths	Pentadactyl limb
3	Power point and scaffolded written exercise	Trilobites
4	Power point and scaffolded written exercise	Pentadactyl limb

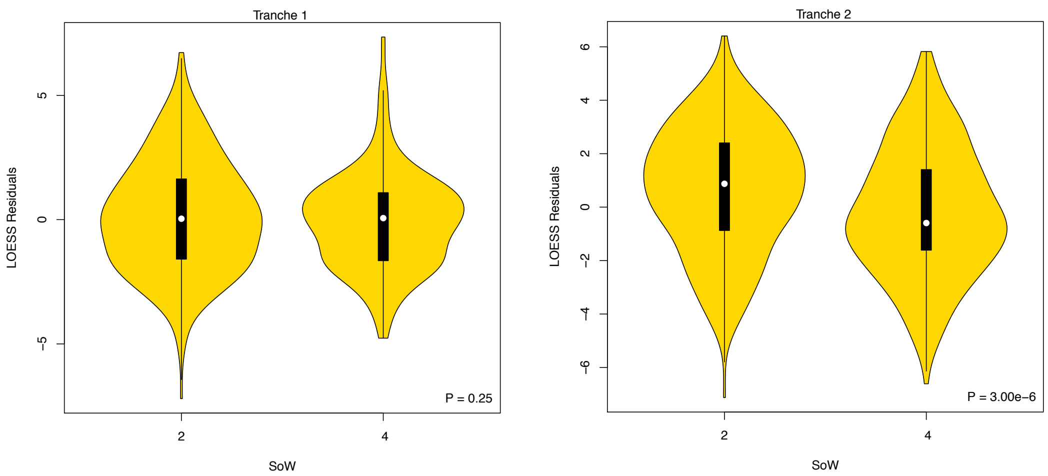
Note: Highlighted rows signifying most effective Schemes of Work and the paired lesson activities contained within them.

Figure 3.15 Violin plots showing the comparison of adjusted LOESS residuals for Schemes of work (SoW) 1 and 3 in both tranches.



Notes: Tranche 1, w = 13545, P = 3.03x10^{-9}, Wilcoxon rank sum test, Cliff's d = -0.34 (moderate effect size); Tranche 2, w = 50198, P = 1.20x10^{-2}, Wilcoxon rank sum test, Cliff's d = -0.14 (negligible effect size).

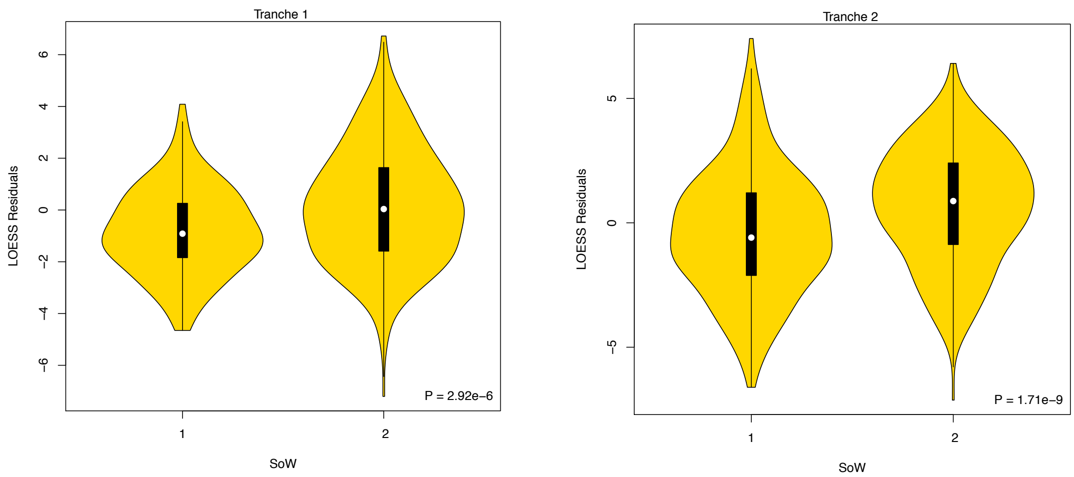
Figure 3.16 Violin plots showing the comparison of adjusted LOESS residuals for Schemes of Work (SoW) 2 and 4 in both tranches



Notes: Tranche 1, w = 43403, P = 0.25, Wilcoxon rank sum test, Cliff's d = 0.05 (negligible effect size); Tranche 2, w = 56348, P = 3.00x10^{-6}, Wilcoxon rank sum test, Cliff's d = 0.22 (small effect size).

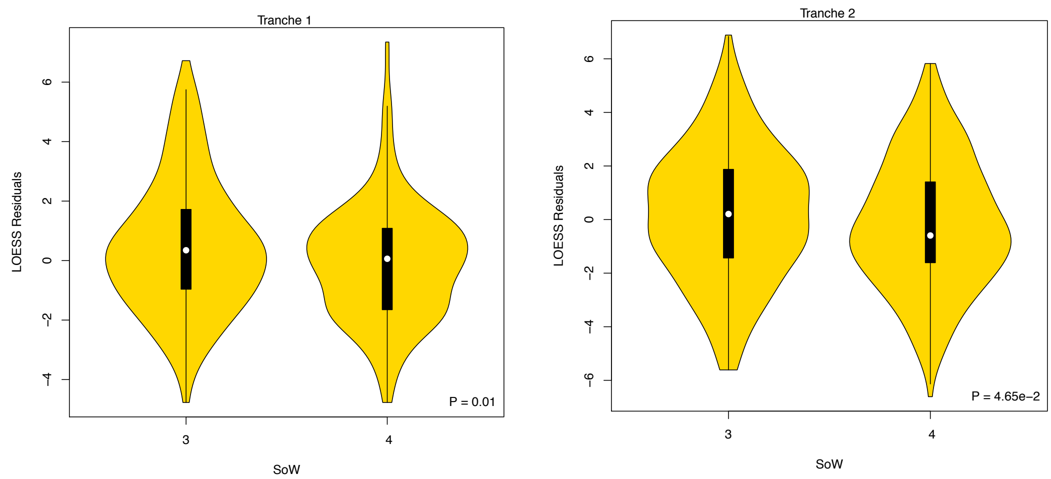
The relationship from the opposite direction can also be considered to ascertain the relative effectiveness of the different homology/common ancestry tasks in lesson 4 preceded by one of the two different moth activities by comparing the last two pairs of different Schemes of Work (SoW); 1 and 2 and 3 and 4. Scheme of Work 2 was found to be significantly more effective than Scheme of Work 1 in both tranches (see Fig. 3.17) with a combined P = 8.04×10^{-11} (Fisher's test) suggesting that the pentadactyl limb task was more effective than studying Trilobites if preceded by hunting paper moth. However, when Schemes of Work 3 and 4 were compared it was found that studying Trilobites was significantly more effective in both tranches than the pentadactyl limb if preceded by the power point moth activity (see Fig. 3.18), with a combined P = 4.95×10^{-3} (Fisher's test). These results suggest positive and significant interactions between the homology/common ancestry tasks in lesson 4 preceded by the different moth activities in lesson 2. Studying Trilobites preceded by completing the power point moth activity or conversely studying the pentadactyl limb preceded by completing the hunting moths activity were most effective in improving student performance. This reciprocal result reinforces the previous Scheme of Work pair analyses confirming that Schemes of Work 2 and 3 were relatively more effective than 1 and 4 due to unexpected positive interactions between the main activities completed in lessons 2 and 4.

Figure 3.17 Violin plots showing the comparison of adjusted LOESS residuals for Schemes of Work (SoW) *1 and 2 in both tranches*



Notes: Tranche 1, w = 19952, P = 2.92×10^{-6}, Wilcoxon rank sum test, Cliff's d = -0.26 (small effect size); Tranche 2, w = 44606, P = 1.71×10^{-9}, Wilcoxon rank sum test, Cliff's d = -0.27 (small effect size).

Figure 3.18 Violin plots showing the comparison of adjusted LOESS residuals for Schemes of Work (SoW) 3 and 4 in both tranches



Notes: Tranche 1, w = 35609, P = 0.01, Wilcoxon rank sum test, Cliff's d = 0.13 (negligible effect size); Tranche 2, w = 48860, P = 4.45x10^{-2}, Wilcoxon rank sum test, Cliff's d = 0.09 (negligible effect size).

3.9 There was strong teacher endorsement of all the main activities

Although this study was designed to determine the relative effectiveness of the 4 different main activities this outcome was not possible due to the unexpected interaction between main activity pairing in lessons 2 and 4. However, the relative effectiveness of the Schemes of Work can be explored from a new perspective by considering post teaching feedback obtained from participating teachers.

First, the feedback on lesson 2 was considered, in which the students explored natural selection in Peppered moths in one of two different ways: either by actually 'hunting' paper moths in a more student centric active manner or by completing a scaffolded written activity after watching and discussing a power point presentation in a more teacher centred approach.

The 'hunting' moths activity involved a differentiated practical investigation in which newspaper and white paper moths were 'hunted' by the students using forceps 'beaks' against different backgrounds followed by some questions to consolidate. This activity was used to introduce the idea of selective predation producing differential survival and reproduction leading to changes in trait frequency in the moth population in subsequent generations. The Peppered moth was chosen as a realist example that the students could relate to:

"They did love the moth hunting activity, I think it was a great example, because sometimes I find in scientific schemes and things, the animals aren't very child friendly. They get moths. They get birds. That's it. It's real for them."

Teachers reported that the although the activity took longer than expected to complete due the additional organisation needed to select the groups and distribute the resources the practical was *worthwhile, as it demonstrated the concept of natural selection very well.*

"The children found this really interesting, really interesting. They loved doing the activity, it took a long time to do, so it took most of the afternoon. We did it as a whole class activity, I had mixed year groups together so the logistics took a bit of organising. It did take a long time to do but it ensued the less able and the younger ones were still able to fill in the sheets and participate."

"It [the moths] took a lot longer to do than I thought it would as they found it difficult to understand the instructions. But I think that was me rather than the kids. I did it to two classes and I was much better at it the second time, obviously I got it then and was better at explaining what to do."

"They saw the concept very clearly, I think it was very visual, which was really good and they, did it without trying, they were absolutely picking up the ones that, you would expect, the ones that really stood out against the background"

Once the small groups had been organised the students were able to just get on with the practical allowing the teacher to act as a facilitator drawing out extra details and developing their understanding of new concepts.

"We did it as a whole class activity, altogether. I said, "OK, go" and everybody did their things. And we were then sort of able to discuss, what are you finding, what's happening in this group, what's happening in that group as I walked round the room, so we did take a long time to do it. They also got the idea that it was a simulation, and what a simulation was, and that was a completely new concept, completely new to them. They found, not just, not just proving camouflage helps, so the white ones obviously you can't see on the white paper, so they can survive better. They actually found that there were patterns in the size as well. So, my focus had been, OK, well we're going to understand by the end of this session that it's the camouflage that is the key factor in survival and therefore reproduction. But they came out with if they're smaller, they're harder to get, as well."

They also remarked on the practical nature of the activity forming a point of reference underpinning the understanding of the concept of natural selection:

"The one with the moths and forceps was brilliant as well, that was something that then we could keep referring back to, so when we were discussing other things, we said, "Oh do you remember?" So, I think the fact that they'd done it practically really helped kind of embed that understanding about that one."

The alternative Peppered moth activity involved a step by step power point presentation explaining the process of natural selection with an embedded an interactive white board class simulation, followed by a scaffolded differentiated story board written exercise. Although less teachers commented on this activity than the previous one the majority of those who did comment reported that their students enjoyed the activity and thought the differentiated scaffolded writing frame was a successful way of describing the process.

"The white board moth story, they loved that as well. It helped really visualise it, you know, seeing the different environments that the Peppered moths were in. Yes, I think they did, and they did some nice write-ups on that as well, so that was one of the really good lessons that we did actually."

"We put the frame up and then they used that frame then to write the second one. Through the use of the scaffold, were they all able to access that to a lesser or greater extent. Yeah, it definitely helped. I think they would have struggled. Well, quite a lot of them would have struggled if they didn't have that."

However, one teacher felt that this activity was not successful in their class.

"We found it quite complicated as well for them. They have to have very, very visual stuff or hands on things and that was a little bit, it just wasn't that interesting."

These mainly positive comments demonstrate that teachers endorsed the use of both different moth activities as a method of introducing natural selection in primary age students. However, using feedback obtained from the small number of teachers who participated in the study over two consecutive years we can explore individual teacher preference by direct comparison of the two moth activities taught by the same teacher. Again, this feedback suggests that both activities were successful in demonstrating the principle of natural selection and engaging student interest and is consistent with the statistical testing carried out in the previous section. Additionally, the choice of which of the two activities they would use in the future would depend on the ability of their students, the time available and the availability of resources.

“I had the white board moth activity this year and that went really well, I think again it's a really, really brilliant way of showing how things survive, it worked well last time to be honest when we hunted the moths. I know it's the same concept but just done in a slightly different way. Both were really good and the seed has been Scheme of Workn for them for their interest in the future, which I think is really important”.

“Having done both [moth activities], I think my top set really enjoyed hunting the paper moths. I think with my parallel sets [middle ability] and low ability it was much easier to manage the white board one, because just organising their equipment and counting out things takes so much time that we were getting to the point of no return and losing things. I think it was it was much easier to use the computer with the write up sheet with my parallel groups”.

If attention is now turned to the concept of common ancestry by studying homology in lesson 4, the data demonstrates that the teachers felt that the pentadactyl limb activity was an interesting way to introduce the concept of common ancestry through the study of homologous bones within the forelimbs of different mammals. Teachers in this study reported that their students enjoyed the activity and that it really helped them to visualise our links to other mammals and shared common ancestry.

“They really enjoyed doing that [limb activity] and seeing the bones and I think again the visual of actually making them was great.”

“They liked that [limb activity]. I think they found it really interesting to see how, you know, how the bones had adapted and changed their positions all the way through from the sheet and colouring in. They liked that they could sort of see the horse’s hoof and all the rest of it, and the salt dough, they liked the salt dough.”

“We had the pentadactyl limbs and we made salt dough life-size versions of them. They got into little groups and they made full limbs and everything, which was really interesting. They were able to label all of them. Again, it was one of those where they saw it, and then they could see the adaptations and that reinforced the having one single ancestor and how that all worked through. That was really interesting, for them. It was quite cool for me as

well. There were a few ‘penny drop’ moments throughout the room. Then they were like, oh yeah, it's the same.”

“They found that [the limbs] really interesting actually looking at how all of these different animals, skeletons had the same bones in and identifying the bones in the different animals and trying to work out which animals they were. They learnt a lot in that lesson and could understand how we share the same common ancestor.”

Teacher preconceptions over concerns that the activity seemed too complicated and difficult for the students to complete before teaching it were allayed:

“*They had the limb. It was amazing actually. I thought they’d find it more difficult than they did. There was very few that needed me to go over and really point out what was going on. Yeah, they really picked up on it. When you look at that task, you think they’re never going to be able to do that because it looks too complicated. But that’s what I thought and they just sort of got their heads down and did it.”*

The teachers also felt that the Trilobite activity in lesson 4 was an interesting way to study how adaptations can aid an organism’s survival in different habitats. They reported that the students enjoyed the practical activity and that it generated some valuable discussions furthering their understanding.

“They loved that one [Trilobites]. The salt dough, that was brilliant, and showing the adaptions. They had to explain their choices. They took the different sections and they had to show adaptions on those sections, then describe them and because of all that learning that went before, they could easily justify their adaptions and why they would possibly work. The range of discussions you’d get within a group were just amazing.”

“Oh, yes, the salt dough trilobites. They very much enjoyed those. We had some very interesting looking things and quite a lot of flour all over the floor. They loved that. A nice practical thing. They were saying things like “Am I supposed to have that”? They look over and think, mine doesn't look like that. Yours won't because it lives somewhere else, and they're like, oh good. Actually, this type of talk and sharing their ideas helps them have a better understanding of their own knowledge.”

“They enjoyed that [the Trilobite activity] At first, they were just loving the dough, it was a great activity. And then they had to talk about their species, how it was adapted, where it was living. Whether it would survive or not and why.”

“I thought the Trilobites particularly were a massive success. They absolutely loved it, because I think it brought to life what's a very difficult subject, I mean obviously you can look at fossils and everything else. But I think making something of their own that looks like something from millions of years ago was just brilliant. The class was able to choose their own adaptions, and so they talked about their trilobite and where it would live and how it would survive and all of those sorts of things. I think it was very, very successful drawing things together at the end of the topic.”

However, none of the teachers commented on the intended link to common ancestry through the study of homologous body plans in *Trilobite sp*. unlike the comments obtained from the pentadactyl limb activity. Additionally, some of the teachers did not seem to particularly like this activity as it was perceived to be boring and out of context with the rest of the sequence of lessons:

“We got trilobites, and the kids said, ”Oh, not more paper.” The colouring in exercise really didn’t work as they would have preferred to have fossils on the table, magnifying glasses and the chrome books and identified what those fossils were and looked to see what they related to now. So again, they wanted to be more hands on.“

Interestingly, when repeating teachers were asked for their preference between the main activities for lesson 4 (either pentadactyl limb or Trilobites) opinions were mixed and depended on class interests and student ability.

“We enjoyed the trilobites more than the limb from last year. It's funny isn't it? I thought the bones would be more relevant to most children but this group were very interested in fossils, a lot of them brought in their own fossils to show. So, I think that's maybe why it appealed to them, I'm not really sure, but possibly that's why it appealed to them so much.“

“We loved them [the Trilobites]. They really enjoyed it but it wasn’t that different from the previous one [limb] which was also really good. I was a bit concerned about it [using this with my low ability group] but once we did one or two together they sort of flew a bit and actually they enjoyed doing it on their own, as I think they felt it was more challenging than [the work] they normally have but more manageable than limbs which were more

complicated. They seemed quite chuffed, like we're doing this bit of work on our own and we can explain what we're doing."

These positive endorsements of the two different activities for both lessons confirm and reinforce the findings of the quantitative analyses; that although it was not possible to determine the best teaching intervention all four different activities were effective and successful in improving student understanding natural selection and evolution.

3.10 The salt dough modelling activity received mixed reviews

A small proportion of the teachers did not think that the salt dough modelling activity was appropriate for use in their classes. Interestingly, a variety of different reasons were given by them teachers for not completing the salt dough component of lesson 4:

1. Some teachers held preconceived ideas regarding the appropriateness of this activity as a learning experience for this age group:

"We didn't actually do the salt dough because we thought it was too babyish"

2. Whilst other teachers thought it would be too messy, take too long or lacked the confidence to use it in class:

"I didn't do the salt dough as it was going to be too messy and I think a little bit of me was kind of dreading the prep beforehand. I think second time round with confidence of lessons, I probably will plan it in, but I was constantly playing with time. I wasn't sure how long each lesson would take and then I wasn't sure whether I had enough time to fit it in".

"So, we did all the other activities, we watched the videos about the evolution of whatever, did the colouring in, but we didn't do the salt dough bit. We didn't have time to do everything, but we did everything else of it, and enough for them to get the learning from it. Next time I would obviously do the salt dough bit, so we'd have to make a bit more time. It wasn't a choice not to do it because I didn't think they'd like it, because they would, they like that kind of stuff. It was purely time".

However, when the teachers overcame their reservations and used the dough in their classes they realised their misgivings were unfounded.

"They absolutely loved the salt dough and I was a bit nervous actually because sometimes I'm quite good academically but I'm not very good practically and I was like, urgh, I've never made salt dough and this could go wrong".

"You gave me a recipe and actually I thought I've got so much to do, and it's over lunch time so I just got a couple of my kids in and said here's a recipe make it. They did it. Obviously, everyone wanted to join in with making the salt dough and they absolutely loved it. The Trilobites looked absolutely fantastic and they took them home."

Other teachers went on to validate the use of modelling with salt dough as a useful practical tool to consolidate student understanding, as well as appreciating that could be applied more widely in different contexts.

"No, I've never used salt dough before in years five and six, so that's definitely something we'll be doing again. And for some of them, having that practical hands on experience makes such a big difference for them, particularly boys".

"I would absolutely use it again in a cross curricular way with art and craft. I have to confess what with the new curriculum and getting to grips with SATS and everything this year, we've not been that good at doing so having the opportunity to do that was brilliant for them".

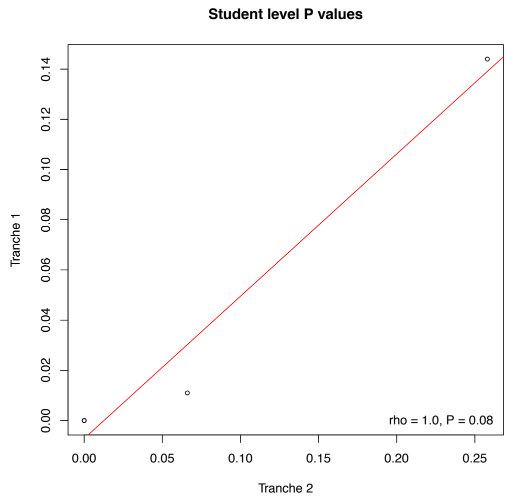
"I've got more confidence with it now so I'd definitely use it in other areas. Yes, and they did that really, really well. Because they were doing salt dough and they had the sheets, they were able to really look at the trilobites in a lot of detail to see how they were adapted and they realised some of the mistakes they made as they went along. They loved it because we're such a writing heavy year with the new writing moderation, just to get them to do something in an afternoon that's a bit more relaxed, just fun they'll remember that more than they will do a big write up."

3.11 Evidence of reproducibility of results in student level analyses

If the student level LOESS residual results show reproducibility between data sets it is expected that the values of P taken from the individual equivalent statistical tests would show significant positive correlations. A non-significant strong positive correlation between P values ($\rho = 1.00$, $P = 0.08$, Spearman's rank correlation coefficient, $n = 4$) was obtained from the four individual student level tests carried out to identify possible predictors of performance (see

Fig. 3.19). The strength of the correlation in this very small sample size (n=4) providing evidence of repeatability at this level of analysis.

Figure 3.19 Scatterplots with lines of regression showing correlation between P values from individual statistical tests in both tranches at the student level of analyses



3.12 Chapter conclusion

In this chapter the student level analyses from both tranches of data together with qualitative evidence from participating teachers has been presented. Analysis of the data has shown that the teaching interventions significantly improved student performance with some degree of longer-term retention evaluated by using an adapted assessment which was both valid and reliable. The use of LOESS residuals to adjust for pre-score and mitigate the ceiling effect has been justified and these values have been used to identify potential predictors of student performance. In the next chapter student preconceptions will be explored in more detail using pre-test data and how responses to individual multiple-choice questions change after teaching.

Chapter 4 What preconceptions do students have when they enter the classroom and are they resistant to change?

4.1 Chapter overview

In this chapter the preconceptions students bring into the classroom before any formal instruction will be explored by examining pre-test questionnaire responses. The methodology employed is outlined and the data analysed to ascertain how resistant these alternative conceptions are to instruction.

4.2 Summary of alternative conceptions and their impact on learning

4.2.1 Children hold many different preconceptions when they enter the classroom

It has been suggested that before formal instruction students in the classroom are not ‘blank slates’ but instead bring with them their own ideas about the world around them and how it was formed (Driver, Leach, et al., 1994). They have already formed naïve theories, common sense explanations for natural phenomena based on their experiences in the real world (Driver, Leach, et al., 1994). These explanations work well in everyday life even if they are not scientifically accurate e.g. young children think the sun rotates around the earth, the heliocentric model contradicts their own observations and intuition, therefore requiring significant cognitive restructuring to adopt the correct model of our solar system (Sinatra et al., 2008). Some of their ideas may be misconceptions but these are preconceptions not mistakes or false beliefs as they are based on their own experiences rather than teaching (Nelson, 2012). Similarly, students in this study have already formed naïve theories about evolutionary concepts and may hold alternative conceptions which must be addressed by creating new experiences that provide the correct scientific evidence to contradict and correct their preconceptions.

If religious and emotional constraints are ignored (see section 1.3.3) even well motivated students find learning about evolution difficult due to a wide variety of alternative conceptions including; cognitive constraints, Lamarckian thinking, soft inheritance and problems with the understanding of common ancestry, species concept and geological time. Each of these alternative conceptions will now be considered in turn.

Cognitive constraints (teleology, essentialism, intentionality)

Sinatra et al. (2008) identify three main cognitive constraints, potential sources of alternative conceptions being detrimental to the understanding of evolution; essentialism, teleology and intentionality.

Essentialism

Essentialism can be defined as a belief that objects possess a set observable characteristics determined by an immutable underlying essence that can't be seen, but gives the object its identity (Shtulman, 2006; Sinatra et al., 2008). Essentialist thinking has the deepest philosophical roots, forming the foundation of the early biological classification systems in which species were grouped according to unchanging observable features, in which any deviations from the normal typed viewed as either irrelevant (Shtulman & Calabi, 2012) or a developmental mistake (Mayr, 1991a, 1991b). It assumes that members of the same species share the same traits which do not alter (Emmons & Kelemen, 2015; Gelman & Rhodes, 2012) and so dismisses the importance of intra-specific variation making natural selection, speciation and macroevolution seem impossible. Recognition of individual variation being vital to understand adaptation as a population based process (Gregory, 2009; Mayr, 1982). Essentialist thinking is important in children for making judgements on how to categorise new objects and situations based on their prior knowledge, forming reasoning short cuts allowing more effective and efficient learning. Students holding essentialist beliefs fail to appreciate the relevance of variation individuals and the transgenerational nature of evolutionary change. Instead they see evolutionary change as a transformation of a species 'essence' (Mayr, 2001) rather than the differential survival and reproduction of individuals with more advantageous traits within a population over many generations (Shtulman, 2006). This view of evolution can lead to the importance of intra specific variation being undervalued or even ignored all together (Shtulman & Calabi, 2013), with some students believing that all individuals within a population of organisms are the same or nearly identical (Anderson et al., 2002; Greene, 1990; Passmore & Stewart, 2002; Shtulman, 2006), only recognising differences between males and females, and between young and old (Flanagan & Roseman, 2011). Students who appreciate the extent of individual-level variability being more likely to have the correct mechanist grasp of natural selection (Shtulman & Schulz, 2008).

Teleology

Children often attribute evolutionary changes to intent or the agency of individual organisms (Kelemen, 2012; Kelemen & Rosset, 2009; Moore, 2002) based on the assumption that the natural world operates in a deliberate, purpose driven way (Kelemen, 1999a; Poling & Evans, 2002). Children have a natural predisposition to teleological explanations, things being made for an end or purpose e.g. birds have wings because they need them to fly (Jensen & Finley, 1996), rocks are pointy to stop animals sitting on them (Kelemen, 1999b). These explanations can mislead and confuse students by implying there is no need to identify the causal mechanism to account for the observation (Tamir & Zohar, 1991). This tendency toward teleological explanations runs very deep and persists throughout secondary school (Southerland, Abrams, et al., 2001) and into tertiary education (Kelemen & Rosset, 2009). In fact, it has been argued that teleology is the default mode way of thinking which is, at best, suppressed by scientific instruction rather than supplanted (Gregory, 2009). A related conceptual constraint to teleology is anthropomorphism, in which human-like conscious intent is attributed either to the objects of natural selection or to the process itself, the forces of nature/evolution transforming individuals in a goal directed way (Urquiza-Haas & Kotrschal, 2015). Anthropomorphic misconceptions can be classified as being either internal (assigning adaptive change to the intentional actions of organisms) or external (regarding the process of natural selection or 'nature' as a conscious agent (Kampourakis & Zogza, 2008). Internal anthropomorphism or 'intentionality' is intimately tied to the misconception that individual organisms evolve in response to challenges imposed by the environment.

Intentionality

Intentionality, can also be assigned to events not only being purposeful but caused by an intelligent agent with a mind of its own (Cid, 2013). Some students assume that evolution is guided externally by a higher power 'theistic evolution' (Scott, 2000) whilst others believe it to be a purposeful and progressive process acting towards the goal of improving each species (Mead & Scott, 2010a; Werth, 2012). Children have a 'centralised mind set' a natural tendency to assume an external directory force when they see order within a system (Resnick, 1996) and should be encouraged to think in a more decentralised way (Petrosino et al., 2015) in order to view natural selection as an emergent process with no single controlling factor or agent (Chi, 2005). There is also widespread poor understanding of randomness and probability (Mead & Scott, 2010b; Wilensky, 1993). Children often assume that phenomena are either entirely random having no pattern at all or entirely deterministic governed by natural laws to create order. For this reason, they tend to misunderstand how emergent processes generate patterns through random interactions.

Lamarckian thinking

Even if students manage to avoid these cognitive constraints, many believe that the process of evolution involves change due to use or disuse of organs, a view developed by Jean-Baptiste Lamarck in 1800. Lamarckian thinking remains common in naïve explanations for why unnecessary organs become vestigial or eventually disappear e.g. the appendix in humans, in students attending middle through to secondary school (Rudolph & Stewart, 1998) and even persisting into adulthood (Bloom & Weisberg, 2007). Modern evolutionary theory recognises several reasons that could account for the loss of complex features (Jeffery, 2008) some of which involve direct natural selection, but none of which is based on disuse alone.

Soft inheritance

The idea that traits acquired during the life time of parents were passed on to their offspring or so called ‘soft’ inheritance remained the common explanation for more than 2000 years and persisted even in Darwin’s life time. Studies have suggested that belief in soft inheritance arises early in childhood as part of a naïve model of heredity (Kargbo et al., 1980; Lawson & Thompson, 1988) and are not overcome by instruction (Jiménez-Aleixandre, 1992; Settlage, 1994) . The intuitive nature of soft inheritance explaining the theory’s persistent appeal in the scientific community and its resistance to correction in the classroom. Understanding of scientifically accepted (hard) inheritance is incompatible with belief in soft inheritance, therefore students must be persuaded to abandon this alternative conception in order to understand the origin of new variation and its relevance to survival in a given environment.

Species concept

Students also hold several related alternative conceptions on the nature of what a species is and how adaption occurs within a species, with some believing that adaptation occurs simultaneously in all individuals of a given species (Bishop & Anderson, 1990; Halldén, 1988). There is also confusion between what occurs to individual organisms and at the species level (Jensen & Finley, 1996), with problems with understanding the equilibrium processes of evolution. Students tend to view evolution as event-like, as a competition between individuals for specific resources rather than as an equilibrium process involving the accumulated effects of small changes across many members of a species (Ferrari & Chi, 1998).

Geological time

Another major source of alternative conceptions is the understanding of geological time with students possessing scant knowledge on the age of the earth (Jensen & Finley, 1996). Darwin (2004, p. 499) recognized that the issue of 'deep time' was a barrier for the understanding and acceptance of his theory, which was confirmed by Dodick and Orion (2003) who went further to suggest that an understanding of deep time is essential to understand evolution. Children are able to relate to time frames of their own lives and their family but geological time requires a new way of thinking (Fail, 2008), as they must to be able to sequence events in temporal order and appreciate the vast duration of time (Cheek, 2013). However, children do not have a clear understanding of 'millions of years' but then neither do adults (Catley & Novick, 2008). Related to this is the failure to appreciate that 99.5% of all evolutionary lines that once existed on earth have become extinct (Mayr, 1997) with some believing that all species originated at the same time and still exist today (Flanagan & Roseman, 2011).

Common ancestry

The concept of common ancestry is also widely misunderstood, Bizzo (1994) finding that 16 to 17 year-old Brazilian students believed that plants and animals did not share a common ancestor in his small study (n = 11). One of the most common alternative conceptions is that only similar species can share a common ancestor, whilst species with no apparent, obvious or superficial similarities cannot share a common ancestor (Hagay, 2005; Poling & Evans, 2004b; Shtulman, 2006). Some even doubt that humans share a common ancestor with any other life form, Miller et al. (2006) finding that 60% of Americans believe that humans were specifically created and did not evolve from earlier life forms. However, Dougherty (2011) asserts that teaching the concept of common ancestry between humans and other groups of animals can decrease rejection of evolution by students whilst Meikle and Scott (2010) have found using the concept of 'cousins' helpful in overcoming the alternative conception that humans evolved from apes.

4.2.2 Alternative conceptions originating in the classroom

As discussed previously, the formation of alternative conceptions can be attributed to children's naïve theories generated from their everyday experiences and their language usage. These conceptions are resistant to change and can seriously interfere with learning as they are formed prior to instruction. However, there are additional sources of alternative conceptions that can be attributed to difficulties within the classroom; incorrect concepts could be formed during the lesson due to lack of understanding or misunderstanding and the transmission of alternative conceptions from the teacher through factually incorrect or inaccurate teaching (Alters & Nelson, 2002; Yip, 1998). Teachers lacking adequate subject knowledge, especially in evolution are likely to hold content specific alternative conceptions (Kikas, 2004). They may then go onto teach these alternative conceptions to their students compounding the problem further (Abrie, 2010; Jarvis et al., 2003) by impeding the accurate scientific conceptual development of their students (Crawford et al., 2005; Kikas, 2004). As there is evidence suggesting that teachers pass on their evolutionary alternative conceptions to their students (Fisher, 2004; Wood-Robinson, 1994) , this project sought to reduce the occurrence of alternative conceptions and to minimize the acquisition of these erroneous ‘taught and learnt’ (Alters & Nelson, 2002) conceptions through the use of carefully designed teaching interventions and teacher training.

Studies show that many secondary school biology teachers of all experience levels hold alternative conceptions of various biological concepts and that these could be passed on to their students through inaccurate teaching or the uncritical use of resources (Abrie, 2010; Sanders et al., 1993; Yip, 1998). The way in which knowledge is presented in class by the teacher is vitally important (Kikas, 2004). Analogy is frequently used to help students relate new information into their prior knowledge (Taylor & Coll, 1997). However, overgeneralization on the basis of analogy can sometimes give rise to alternative conceptions (Gentner, 1983). The correct interpretation of scientific terms and the misunderstanding caused by confusing everyday meanings is also another source of alternative conceptions (Williams, 2013) as are the diagrams or models chosen to facilitate a better understanding of the topic especially if they are not properly constructed (Mayer, 2002). Additionally, the way in which science is portrayed in class is also very important. Many teachers think that science is based around a known set of facts or laws and so may not present it as a dynamic body of knowledge which is constantly being reviewed and amended as new evidence is discovered (Driver et al., 1996).

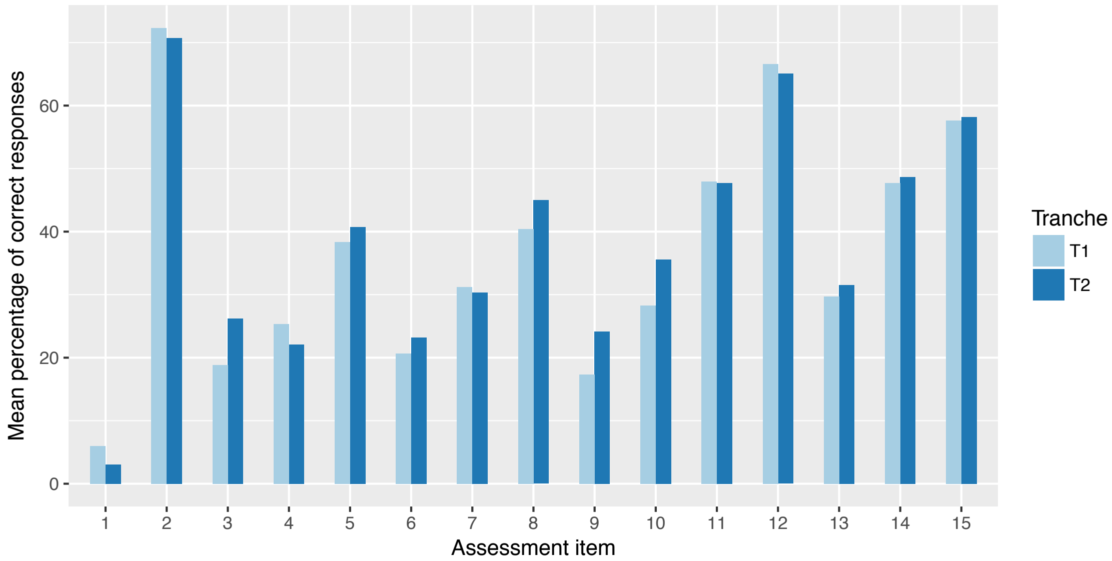
4.3 Results

4.3.1 There was a significant positive correlation between student pre-teaching preconceptions in both tranches of data

First, a visual inspection of the mean percentage of correct responses obtained in the pre-teaching questionnaires in the two tranches of data was carried out; Fig.5.1. There was a good agreement (< 5% difference) in 12 of the 15 items with only items 3, 9 and 10 varying by

> 5% (7.41%, 6.76 % and 7.25% respectively). There was a wide range in the % of correct responses with item 1 being the least well answered and item 2 being the highest scoring item in both tranches.

Figure 4.1 Bar plot of the mean percentage of correct responses for each assessment item in both tranches of data.

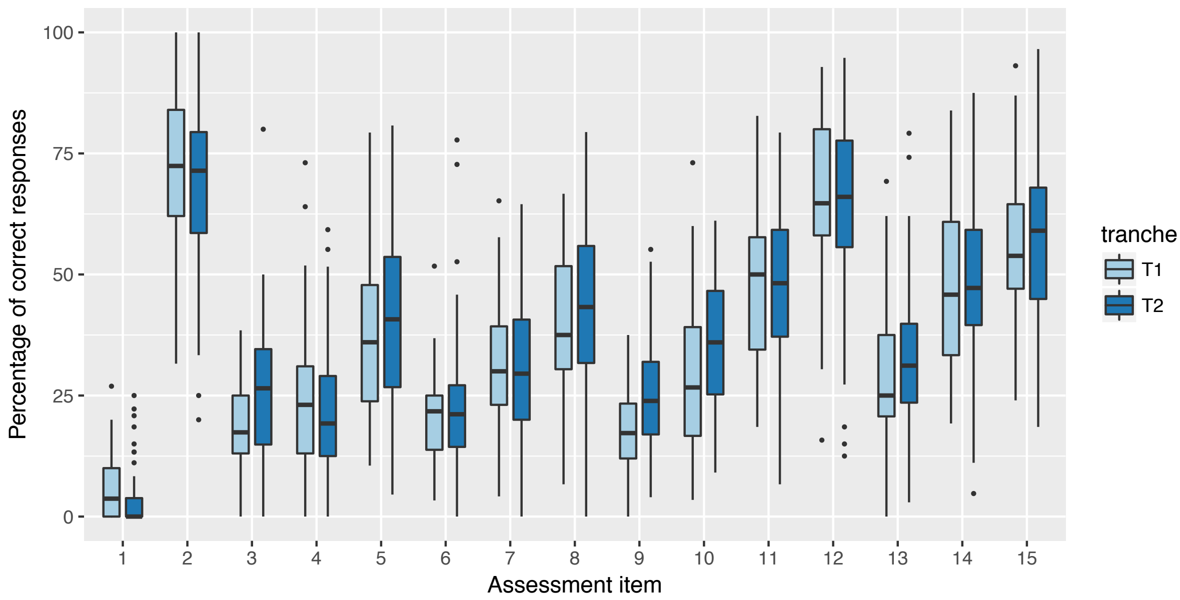


Notes: n = 45 classes in tranche 1 (T1) and n = 58 classes in tranche 2 (T2).

The next step was to correlate the mean percentage of correct responses for each assessment item as a measure of test-retest reliability, another measure of internal reliability. A partial correlation (Pearson's) was carried out to determine the relationship between the results from the two tranches of data whilst controlling for assessment item. There was a strong, positive correlation which was statistically significant between the results of the two tranches of data ($r = 0.98$, $n = 15$, $P = 8.18 \times 10^{-10}$, Pearson's partial correlation), which indicated excellent test-retest reliability for this assessment instrument.

Finally, the difference in mean class pre-test percentage of correct responses between the tranches (Fig. 4.2) was considered and an unpaired analyses for statistical significance carried out. In addition to items 3 ($P = 0.0085$, Wilcoxon test), 9 ($P = 9.4 \times 10^{-4}$, Welch's t test) and 10 ($P = 0.022$, Welch's t test), assessment item 1 ($P = 0.026$ Wilcoxon test) was also significantly different in the two tranches of data. The discrepancy in the findings for assessment item 1 probably being the result of multiple outliers in the data from Tranche 2.

Figure 4.2 Grouped boxplot comparing the percentage of correct responses for both tranches of data startified by assessment item

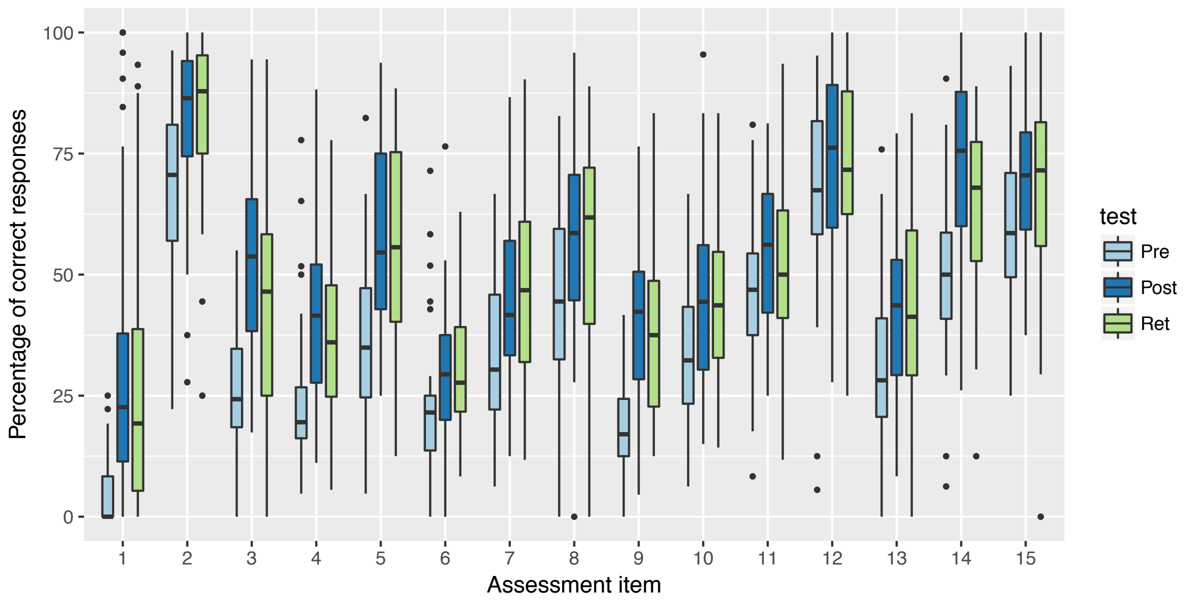


Notes: n = 45 classes in tranche 1(T1) and n = 58 classes in tranche 2 (T2).

4.3.2 Instruction altered student preconceptions.

If the responses from students who completed all of the three assessments are considered in order to compare how their initial preconceptions changed after instruction and whether these changes were maintained over time, much smaller samples are left (n = 320 in tranche 1 and n = 523 in tranche 2). For this reason, the results from the two tranches were combined into 1 set (n = 843 in 40 separate classes). The mean percentage of correct answers for each of the 40 classes was compared for the matched pre, post and retention tests for each assessment item in turn; all assessment items achieving higher post and retention test scores compared to their corresponding pre-test score (Fig. 4.3). Analyses of repeated measures were then carried out on each assessment item (ANOVA for items 3, 5, 7, 8, 9, 10, 11 and 13 as their data were normally distributed and Friedman tests for the other items which were nonparametric) this was followed by pairwise comparison using Nemenyi multiple comparison test (Tukey) and then correction for multiple tests (Holm or Bonferroni).The results from the individual assessment items were then grouped into three main themes; common ancestry and difference between species (items 13, 2, 3 and 1), natural selection (items 8, 4, 5, 6, 9 and 7) as well as fossils, extinction and geological time (items 10, 12, 14, 11 and 15).

Figure 4.3 Grouped box plot comparing the mean percentage of correct answers for students completing all three tests stratified by assessment item



4.3.2.1 Common ancestry and differences between species

This section explores the preconceptions that exist within 4 assessment items investigating the similarities and differences between individuals together with the concept of common ancestry; Table 4.1. Each assessment item is considered in turn.

Table 4.1 Summary of assessment items covering common ancestry and differences between species.

Assessment item	Assessment item focus
13	Similarities and differences between extant species
2	Similarities and differences between extant and extinct species and whether they can share common ancestry
3	Common ancestry in animals and plants
1	Common ancestry in animals including humans and chimps

In order for students to understand the concept of common ancestry they must be able to recognise that there are similarities between members of different species. Item 13 assessed whether students could recognize that there were similarities (and differences) between two diverse species (oak trees and lizards). The correct answer was A, that the two species show both similarities and differences, with response C, showing that they had no similarities only differences based on the common essentialist alternative conception that species with no apparent, obvious, or superficial similarities have no similarities at all (Shtulman, 2006). Pre-instruction the most frequent response was C showing that this alternative conception was held by a large proportion of the students (36.89%) with less students giving the correct response (31.44%). This assessment item showed a significant overall difference in test values post instruction ($F = 12.44$, $P = 2.05 \times 10^{-5}$ repeated measures one-way ANOVA; Table 4.2). Post instruction there was a highly significant conversion to the correct response from all distractors in both the post ($P = 2.2 \times 10^{-4}$) and retention ($P = 9.5 \times 10^{-6}$) tests, the correct response (A) becoming the highest scoring answer in both. The frequency of response C decreased post instruction in consecutive tests suggesting that instruction was successful correcting this alternative conception; that there no similarities between diverse species. In this item there was no evidence of waning from the correct response in the retention test, with a slight increase from the post to retention and no significant difference ($P=0.43$) between the post and retention scores.

Table 4.2 Analysis of assessment item 13. Data shown as mean percentage.

Item 13 Response	Pre-test %	Post-test %	Δ post-pre %	Retention test %	Δ ret-pre %	Δ post-ret %
A	31.44	42.35	10.91****	46.74	15.30****	-4.39
B	8.66	5.81	-2.85	7.59	-1.07	-1.78
C	36.89	34.64	-2.25	29.54	-7.35	5.10
D	21.12	16.96	-4.15	15.78	-5.34	1.19
U	0.59	0.12	-0.47	0.36	-0.24	-0.24
NA	1.30	0.12	-1.19	0.00	-1.30	0.12

*Notes: highlighted row showing correct response, orange rows showing attractive distractors. **** = P ≤ 0.0001*

The next step is to ascertain whether students could recognize whether extinct and extant species could share a common ancestor that lived a long time ago even if they had few similarities using assessment item 2. The correct answer was B, with the attractive distractor A, stating that a living species and an extinct one can share a common ancestor but only if they have lots of similarities. This was based on the common alternative conception that only similar species can share a common ancestor, whilst species with no apparent, obvious or superficial similarities cannot share a common ancestor (Hagay, 2005; Poling & Evans, 2004b; Shtulman, 2006). Pre-instruction the correct response was the highest scoring option (69.51%), with response A being the next frequent. This assessment item was showed a significant overall difference in test values post instruction ($\chi^2 = 33.56$, $P = 5.15 \times 10^{-8}$, Friedman test; Table 4.3) and remained as the highest scoring item. Post instruction there was a highly significant conversion to the correct response from all distractors in both the post ($P = 2.3x10^{-5}$, post hoc Friedman Nemenyi test) and retention ($P = 1.1x10^{-6}$, post hoc Friedman Nemenyi test) tests. There was no evidence of waning from the correct response in the post-test to the retention test with a slight gain (1.30%), whilst the frequency of response A continued to decrease in consecutive tests. There was no significant difference between post and retention scores (P = 0.8, post hoc Friedman Nemenyi test) suggesting that instruction was successful in correcting the alternative conception that extinct and extant species can share a common ancestor even if they only share a few similarities and that this understanding was retained.

Table 4.3 Analysis of assessment item 2. Data shown as mean percentage.

Item 2 Response	Pre-test %	Post-test %	Δ post - pre %	Retention %	Δ ret-pre %	Δ post-ret %
A	13.88	10.08	-3.80	8.42	-5.46	1.66
B	69.51	83.16	13.64****	84.46	14.95****	-1.30
C	6.64	3.56	-3.08	3.56	-3.08	0.00
D	8.19	3.08	-5.10	3.44	-4.74	-0.36
U	0.47	0.00	-0.47	0.12	-0.36	-0.12
NA	1.30	0.12	-1.19	0.00	-1.30	0.12

*Notes; green highlighted row showing correct response, orange row showing attractive distractors. **** = P ≤ 0.0001 using Freidman test: post hoc Friedman Nemenyi test.*

Next, assessment item 3 which covers the shared common ancestry of all plants and animals is considered. The other responses were all attractive distractors based on common alternative conceptions. Response A was the correct answer in which all plants and animals share a common ancestor, with responses B and C both representing the alternative conception that animals and plants do not share a common ancestor (Bizzo, 1994) and D representing the alternative conception that members of different species do not share a common ancestor (Poling & Evans, 2004b; Shtulman, 2006). Pre-instruction option C that only animals share a common ancestor with each other was the was the most frequent response (28.83%) followed by the correct response A (26.69), interestingly option B in which only plants share a common ancestor was less frequent (21.83%). Post instruction this item showed the most significant overall difference in test values ($F = 33.07$, $P = 3.9x10^{-11}$ repeated measures one-way ANOVA; Table 4.4). There was a highly significant conversion to the correct response from all distractors in both the post ($P = 3.3x10^{-8}$) and retention ($P = 4.2x10^{-7}$) tests, the correct response (A) becoming the highest scoring answer in both. There was some degree of waning from the correct response in the post-test to the retention test, however this was not a significant (P = 0.06). The results of this assessment instrument suggest that instruction was successful in both improving and retaining the correct understanding of this concept. Additionally, there was a sequential decrease in the percentage of students opting for response D suggesting instruction was particularly successful in addressing the alternative conception that members of different species do not share a common ancestor.

Table 4.4 Analysis of assessment item 3. Data shown as mean percentage.

Item 3 Response	Pre-test %	Post-test %	Δ post-pre %	Retention test %	Δ ret-pre %	Δ post-ret %
A	26.69	55.28	28.59****	46.26	19.57****	9.02
B	21.83	14.71	-7.12	19.45	-2.37	-4.74
C	28.83	20.17	-8.66	25.39	-3.44	-5.22
D	19.93	9.73	-10.20	8.54	-11.39	1.19
U	0.83	0.12	-0.71	0.36	-0.47	-0.24
NA	1.90	0.00	-1.90	0.00	-1.90	0.00

*Notes: green highlighted row showing correct response, orange rows showing attractive distractors. **** = P ≤ 0.0001*

Assessment item 1 was the final question of this group and proved to be the most challenging for the students, achieving the lowest score on all three tests. The correct response was D in which chimpanzees, humans, zebras, and worms all share a common ancestor, with responses A and C acting as attractive distractors. Response A was based on the common alternative conception that only similar species (humans and chimpanzees) can share a common ancestor, whilst (lions and worms) species with no obvious similarities cannot (Hagay, 2005; Poling & Evans, 2004b; Shtulman, 2006), whilst response C was based another common alternative conception that members of different species do not share a common ancestor (Poling & Evans, 2004b; Shtulman, 2006). This item showed a highly significant overall difference in test values post instruction (χ^2 = 33.99, P = 4.17x10^{-8}, Friedman test; Table 4.5). Pre-instruction option A was by far the most frequent response (84.70%) showing a strong preference for the alternative conception of only closely related species sharing a common ancestor, whilst the correct option frequency was very low (4.39%). Post instruction there was a highly significant conversion to the correct response from all distractors in both the post (P = 1.9x10^{-6}, post hoc Friedman Nemenyi test) and retention (P = 6.4x10^{-5}, post hoc Friedman Nemenyi test) tests, although response A remained the highest scoring answer. There was some degree of waning from the correct response in the retention back to response

A, however this was not a significant (P = 0.75, post hoc Friedman Nemenyi test) suggesting that instruction was successful in correcting preconceptions in some students to the accepted scientific understanding in the longer term. However, it must be noted that the alternative conception represented by response A was both prevalent and persistent remaining as the highest scoring response in all tests, suggesting a deeply held belief that humans and chimpanzees can share a common ancestor as they are so similar. This finding seemingly contradicts the results from assessment item 3 in which students show a much better understanding that all animals and plants share a common ancestor and could be result of the human related context of the assessment item reinforcing this alternative conception.

Table 4.5 Analysis of assessment item 1. Data shown as mean percentage.

Item 1 Response	Pre-test %	Post-test %	Δ post-pre %	Retention %	Δ ret-pre %	Δ post-ret %
A	84.70	60.97	-23.72	66.79	-17.91	-5.81
B	6.76	5.58	-1.19	4.98	-1.78	0.59
C	2.61	1.90	-0.71	0.71	-1.90	1.19
D	4.39	31.44	27.05****	27.40	23.01****	4.03
U	0.59	0.12	-0.47	0.12	-0.47	0.00
NA	0.95	0.00	-0.95	0.00	-0.95	0.00

*Notes: green highlighted row showing correct response, orange rows showing attractive distractors. **** = P ≤ 0.0001 using Freidman test: post hoc Friedman Nemenyi test.*

4.3.2.2 Natural selection

In this section the preconceptions that exist within 6 assessment items investigating the process of natural selection are explored in turn; Table 4.6.

Table 4.6 Summary of assessment items investigating the process of natural selection.

Assessment item	Assessment item focus
8	Inherited traits aiding survival
4	Inherited traits aiding survival and differential reproduction
5	Inherited traits leading to adaptation
6	Change in a species over time
9	Intra-specific variation leading to differential reproduction
7	Mechanism of natural selection

Natural selection can be defined as the process by which random mutational changes are selected for by nature in a consistent, orderly, non-random way. When coupled with descent with modification, natural election can cause a population to evolve for fitness within a given environment over multiple generations.

In order to understand the process natural selection students must overcome essentialist cognitive constraints in order to perceive that variation exists within a species (intraspecific variation) with advantageous traits aiding survival. Assessment item 8 explores the understanding of how inherited traits can aide survival. The correct response was A, that members of the same species can vary in their ability to find food and to reproduce. Response D represented the common essentialist alternative conception that all individuals within a population of organisms are the same, any differences among them being trivial and unimportant (Anderson et al., 2002; Greene, 1990; Passmore & Stewart, 2002; Shtulman, 2006). There was a strong understanding of this concept pre-instruction, with option A being the most frequent response (46.50%) which increased in consecutive tests, showing a significant overall difference in test values post instruction ($F = 6.96$, $P = 1.6\text{x}10^{-3}$ repeated measures one-way ANOVA; Table 4.7). Post instruction there was a significant conversion to the correct response from all distractors in both the post ($P = 0.003$) and retention ($P = 0.003$) tests. There was no evidence of waning from the correct response in the retention test, with a slight increase from the post to retention test (0.47%) with no significant difference ($P = 0.94$) between the post and retention scores. There was also a sequential reduction in the frequency of option D suggesting that instruction was successful in improving the understanding of this concept whilst reducing essentialist conceptions and that this improved understanding was retained.

Table 4.7 Analysis of assessment item 8. Data shown as mean percentage.

Item 8 Response	Pre-test %	Post-test %	Δ post - pre %	Retention %	Δ ret-pre %	Δ post-ret %
A	46.50	57.89	11.39**	58.36	11.86**	-0.47
B	25.15	22.78	-2.37	22.06	-3.08	0.71
C	17.56	12.34	-5.22	13.17	-4.39	-0.83
D	9.13	6.76	-2.37	5.93	-3.20	0.83
U	0.47	0.24	-0.24	0.24	-0.24	0.00
NA	1.19	0.00	-1.19	0.24	-0.95	-0.24

*Notes: green highlighted row showing correct response, orange row showing attractive distractor. ** = P ≤ 0.01.*

Assessment item 4 investigated another aspect of how advantageous traits can aid survival by helping to avoid predation and subsequent differential reproduction. The correct response was option A which stated, that when a new predator is introduced into a population of lizards, the individual lizards with the traits that best help them avoid the predator are more likely to survive and reproduce. All of the other response options were attractive distractors, responses B and C based on anthropomorphic reasoning, that Individual organisms can deliberately try to develop new heritable traits because they need them for survival (Bishop & Anderson 1990; Passmore et al., 2002; Stern & Roseman, 2004), the difference between them being that in option B all lizards would try to develop new traits whereas in option C only some would. Response D was based on essentialist thinking that all individuals within a population of organisms are the same and that any differences among them being trivial and unimportant (Anderson et al., 2002; Greene, 1990; Passmore & Stewart, 2002; Shtulman, 2006). Before instruction the essentalist response D was the most popular however, after instruction there was a signifcant conversion to the correct response A (χ^2 = 23.45, P = 8.09x10^{-6}, Friedman test; Table 4.8). Post instruction there was a highly significant conversion to the correct response A mostly at the expense of option D in both the post (P = 2.3x10^{-5}, post hoc Friedman Nemenyi test) and retention (P = 0.0019, post hoc Friedman Nemenyi test) tests, response A remaining the highest scoring answer in both post instruction tests. There was some degree of waning in the retention test from the correct response back to response A, however this was not significant (P = 0.54, post hoc Friedman Nemenyi test) suggesting that instruction was successful in helping to overcome essentialist preconceptions in some students and their accommodation of the accepted scientific understanding in the longer term. However, it must be noted that the alternative conceptions represented by responses B and C were both prevalent and persistent in all three tests, suggesting that anthropomorphic cognitive constraints are widespread and resistant to change.

Table 4.8 Analysis of assessment item 4. Data shown as mean percentage.

Item 4 Response	Pre-test %	Post-test %	Δ post - pre %	Retention %	Δ ret-pre %	Δ post-ret %
A	24.91	43.77	18.86****	38.32	13.40**	5.46
B	20.17	20.40	0.24	21.83	1.66	-1.42
C	22.30	19.81	-2.49	21.47	-0.83	-1.66
D	30.96	16.01	-14.95	18.03	-12.93	-2.02
U	0.24	0.00	-0.24	0.12	-0.12	-0.12
NA	1.42	0.00	-1.42	0.24	-1.19	-0.24

*Notes: green highlighted row showing correct response, orange rows showing attractive distractors. ** = P < 0.01, **** = P ≤ 0.0001 using Freidman test: post hoc Friedman Nemenyi test.*

Assessment item 5 also investigated another example of how intraspecific variation can aid survival and lead to adaptation in future generations. This item explored student understanding of how the shape of a bird's beak could help it to survive. The correct response was option A, only birds with bigger beaks would be able to get enough food to survive and reproduce, passing on this big beak trait, if only large seeds were available. All other responses were attractive distractors based on common alternative conceptions. Response B was based on Lamarckian thinking, that change occurs in the inherited characteristics of a population of organisms over time because of the use or disuse of a particular characteristic (Bishop & Anderson, 1990), in this case the beaks got bigger because they used them. Whilst response C was based on anthropomorphic thinking, that individual organisms can deliberately develop new heritable traits because they need them for survival (Bishop & Anderson 1990; Passmore et al., 2002; Stern & Roseman, 2004) i.e. birds with small beaks grew them because they wanted them. Finally, response D was based on the alternative conception that change to the characteristics of populations (i.e. the proportion of individuals in the population having certain traits) of organisms is always random and is not influenced by the favourability of that change in a given environment (Flanagan & Roseman, 2011). In this option it was chance that all the beaks got bigger in one generation, so that they could survive and pass on this trait. Pre-instruction the correct response A was the most frequent (36.42%) with option B being the next frequent option. Post instruction this assessment item showed a highly significant overall difference in test values post instruction ($F = 25.24$, $P = 3.5 \times 10^{-9}$ repeated measures one-way ANOVA; Table 4.9). There was a highly significant conversion to the correct response from all distractors in both the post ($P = 9.19 \times 10^{-11}$) and retention ($P = 7.41 \times 10^{-8}$) tests, the correct response (A) remaining as the highest scoring answer. There was some evidence of waning from the correct response between the post and retention tests but this was not significant ($P = 0.25$), suggesting that instruction was successful in improving the understanding of this concept and that this improved understanding was retained. Additionally, options C and D continued to decrease in frequency from the post to retention tests suggesting there was a sustained conversion away from these alternative conceptions. Whereas, option B initially decreased in the post test but then increased to within 1.5% of its pre-test value, suggesting that the Lamarckian thinking represented by this response was a persistent alternative conception which was not fully accommodated by over a quarter of the students post instruction.

Table 4.9 Analysis of assessment item 5. Data shown as mean percentage.

Item 5 Response	Pre-test %	Post-test %	Δ post - pre %	Retention %	Δ ret-pre %	Δ post-ret %
A	36.42	59.07	22.66****	56.23	19.81****	2.85
B	27.64	19.93	-7.71	26.22	-1.42	-6.29
C	13.64	10.08	-3.56	7.47	-6.17	2.61
D	20.88	10.79	-10.08	9.85	-11.03	0.95
U	0.24	0.12	-0.12	0.12	-0.12	0.00
NA	1.19	0.00	-1.19	0.12	-1.07	-0.12

*Notes: green highlighted row showing correct response, orange rows showing attractive distractors, **** = P ≤ 0.0001*

Now that the inheritance of certain traits can affect the survival of members of a species has been established, the understanding of how these traits are passed on to the next generation can be explored. Assessment item 6 investigates the understanding of how a species can change over time, the correct response C being, that individuals born with helpful traits are more likely to survive and pass these traits on to their offspring. Responses A and C both contained elements of soft inheritance in which changes acquired during an individual's life time can be passed on to their offspring. Response A was based on the common alternative conception that changes in a population occur through a gradual change in all members of a population, not from the survival of a few individuals that preferentially reproduce (Anderson et al., 2002; Bishop & Anderson, 1990; Brumby, 1984). Response B was similar but with only some of the individuals surviving rather than all of them. Response D was based on the essentialist alternative conception that the internal chemistry, appearance, and behaviour of a species do not change, even over long periods of time (Flanagan & Roseman, 2011). Pre-instruction, both of the soft inheritance options were more frequent than the correct response. Although this item showed a significant overall difference in test values post instruction (χ^2 = 7.49, P = 0.024, Friedman test; Table 4.10), the post hoc test was inconclusive with no significant difference between the pairs of tests. Although the correct response C increased across successive tests with no evidence of waning, response A remained the highest scoring option with a slight increase in the retention test mainly caused by students switching from alternative conception B. The persistence and prevalence of responses A and B suggesting that the belief in soft inheritance was wide spread and resistant to change even after instruction.

Table 4.10 Analysis of assessment item 6. Data shown as mean percentage.

Item 6 Response	Pre-test %	Post-test %	Δ post - pre %	Retention %	Δ ret-pre %	Δ post-ret %
A	35.82	34.28	-1.54	38.08	2.25	-3.80
B	33.45	32.50	-0.95	28.83	-4.63	3.68
C	23.13	28.83	5.69	29.42	6.29	-0.59
D	6.05	4.15	-1.90	3.44	-2.61	0.71
U	0.59	0.12	-0.47	0.24	-0.36	-0.12
NA	0.95	0.12	-0.83	0.00	-0.95	0.12

Notes: green highlighted row showing correct response, orange rows showing attractive distractors.

Assessment item 9 explored the process of natural selection that individuals of the same species may differ in their inherited traits, and these differences may affect their relative success in survival and reproduction. Option A was the correct response, members of the same species may have inherited different traits and that these differences may change their chance of surviving and reproducing. Responses B and C were attractive distractors, both containing an essentialist element; that members of the same species have the same inherited traits and that any differences are trivial and unimportant. However, they differed in one major way, whilst option B gave an equal chance of surviving and reproducing, option C included an element of soft inheritance, that individuals could pick up new skills during their lifetime which altered their chances of surviving and reproducing. Pre-instruction option C was the most frequent response with over half of the students opting for it (50.18%), with less than a fifth of students (18.86%) selecting A, the correct response. This item showed a highly significant overall difference in test values post instruction ($F = 33.56$, $P = 3.1\times10^{-11}$ repeated measures one-way ANOVA; Table 4.11). Post instruction there was a highly significant conversion to the correct response from all distractors in both the post ($P = 6.6\times10^{-14}$) and retention ($P = 3.5\times10^{-10}$) tests, although response C scored slightly higher in the retention test (0.23%). There was no evidence of waning away from the correct response in the retention test with a slight increase from the post to retention test (4.51%), however this was not a significant ($P = 0.21$) suggesting that instruction was successful in correcting some alternative conceptions to the accepted scientific understanding in the longer term. However, it must be noted that the alternative conception represented by response C was both prevalent and persistent remaining as the highest scoring response in the retention test, confirming that alternative conceptions based on soft inheritance are resistant to change by instruction.

Table 4.11 Analysis of assessment item 9. Data shown as mean percentage.

Item 9 Response	Pre-test %	Post-test %	Δ post - pre %	Retention %	Δ ret-pre %	Δ post-ret %
A	18.86	41.87	23.01****	37.37	18.51****	4.51
B	14.95	10.91	-4.03	13.17	-1.78	-2.25
C	50.18	36.89	-13.29	37.60	-12.57	-0.71
D	13.88	10.20	-3.68	11.15	-2.73	-0.95
U	0.83	0.00	-0.83	0.47	-0.36	-0.47
NA	1.30	0.12	-1.19	0.24	-1.07	-0.12

*Notes: green highlighted row showing correct response, orange rows showing attractive distractors, **** = P ≤ 0.0001*

Assessment item 7 explored student understanding of what is necessary for the process of natural selection to occur. The correct response was D, that traits must be inherited from one generation to the next in order for natural selection to occur, with response C representing the alternative conception that sudden environmental change is required for natural selection to occur (Nehm & Reilly, 2007). Pre-instruction the correct response D was the most frequent (34.40%) with a fairly equal distribution of scores between the other three options. This item showed a significant overall difference in test values post instruction ($F = 8.09$, $P = 6.4x10^{-4}$, repeated measures one-way ANOVA; Table 4.12). There was a significant conversion to the correct response in both the post ($P = 1.2x10^{-3}$) and retention ($P = 1.2x10^{-3}$) tests, the correct response (A) remaining as the highest scoring answer in all tests. The increase in the correct response was obtained by conversion from options A and B rather than from the alternative conception C which remained fairly constant, suggesting that the alternative conception; that sudden environmental change is needed for natural selection to occur is persistent and resistant to change even after instruction. There was some degree of waning from the correct response form post-test to the retention test, however this was not a significant ($P = 0.91$) suggesting that instruction was successful in correcting alternative conceptions to the accepted scientific understanding in the longer term.

Table 4.12 Analysis of assessment item 7. Data shown as mean percentage.

Item 7 Response	Pre-test %	Post-test %	Δ post - pre %	Retention %	Δ ret-pre %	Δ post-ret %
A	20.28	10.56	-9.73	11.39	-8.90	-0.83
B	22.42	18.03	-4.39	19.22	-3.20	-1.19
C	20.28	23.61	3.32	22.18	1.90	1.42
D	34.40	47.09	12.69**	46.50	12.10**	0.59
U	1.07	0.36	-0.71	0.47	-0.59	-0.12
NA	1.54	0.36	-1.19	0.24	-1.30	0.12

*Notes: green highlighted row showing correct response, orange row showing attractive distractor, ** = P ≤ 0.01*

4.3.2.3 Fossils, extinction and geological time

Finally, the preconceptions that exist within 5 assessment items investigating the fossils, extinction and geological time scales are explored; Table 4.13.

Table 4.13 Summary of assessment items covering fossils, extinction and geological time scales

Assessment item	Assessment item focus
10	Nature of fossils
12	Uses of fossils
14	Extant species evolution time frame
11	Extinctions through geological time
15	Changes in the environment over geological time

The first two assessment items in this section concern the nature of what constitutes a fossil and what the study of fossils can tell us. Although the teaching interventions did not directly contain any activities designed to cover this aspect of the national curriculum, reference was made to this content in the starter for lesson 3 and students should have covered fossils in year 3. Assessment item 10 investigated the understanding of what a fossil is, the correct response being C; the term 'fossil' can refer to both a bone in which the original matter has been replaced by rock as well as an impression left by a bone in rock. Less than a third of students (32.50%) selected the correct response before instruction. This item showed a significant overall difference in test values post instruction (F = 7.99, P = 7.0×10^{-4}, repeated measures one-way ANOVA; Table 4.14), with a significant conversion to the correct response in both the post (P = 1.2×10^{-3}) and retention (P = 1.4×10^{-3}) tests. The correct response (C) selected most frequently in both post instruction tests. The increase in the correct response was obtained by conversion from all of the other options which decreased in frequency across successive tests. There was no evidence of waning from the correct response with a further slight increase (1.66%) in the retention test, however, this was not a significant (P = 0.89) suggesting that instruction was successful in correcting alternative conceptions to the accepted scientific understanding in the longer term.

Table 4.14 Analysis of assessment item 10. Data shown as mean percentage.

Item 10 Response	Pre-test %	Post-test %	Δ post - pre %	Retention %	Δ ret-pre %	Δ post-ret %
A	23.13	21.12	-2.02	20.88	-2.25	0.24
B	35.23	30.72	-4.51	29.66	-5.58	1.07
C	32.50	43.77	11.27**	45.43	12.93**	-1.66
D	7.24	4.03	-3.20	3.68	-3.56	0.36
U	0.59	0.12	-0.47	0.36	-0.24	-0.24
NA	1.30	0.24	-1.07	0.00	-1.30	0.24

*Notes: green highlighted row showing correct response, ** = $P \leq 0.01$*

The understanding of what the study of fossils could tell us was investigated by assessment item 12. The correct response was A, that a scientist studying the fossils of an extinct species of fish could discover what anatomical features the extinct species had and could discover similarities and differences between the features of the extinct fish and those of extant fish. Pre-instruction nearly 70% of the students gave the correct response with ~ 12% opting for responses B and C in which only one of the two uses of fossils were given. This was the only assessment item that did not give a significant difference post instruction ($\chi^2 = 5.09$, P = 0.078, Friedman test; Table 4.15), probably due to the very high percentage of correct responses in the pre-test (~70%). There was conversion to the correct response from options B and D in the post test, however, the frequency of option C, that by studying a fossil fish one could discover similarities and differences between it and an extant species remained fairly constant across all of the tests. There was no evidence of significant waning in the retention test with the majority of students understanding how the study of fossils can help us compare features between extinct and extant organisms.

Table 4.15 Analysis of assessment item 12. Data shown as mean percentage.

Item 12 Response	Pre-test %	Post-test %	Δ post - pre %	Retention %	Δ ret-pre %	Δ post-ret %
A	69.87	73.43	3.56	73.31	3.44	0.12
B	12.34	10.56	-1.78	9.49	-2.85	1.07
C	12.22	12.34	0.12	12.34	0.12	0.00
D	3.68	3.32	-0.36	4.27	0.59	-0.95
U	0.95	0.36	-0.59	0.47	-0.47	-0.12
NA	0.95	0.00	-0.95	0.12	-0.83	-0.12

Notes: green highlighted row showing correct response

To explore student understanding of when most extant species evolved assessment item 14 can be examined. The correct response C, that most species living today did not exist at the time life began was given by ~ 50% of students in the pre-test suggesting that this concept was well understood before instruction. Response A was based on the alternative conception that all species began at the same time and still exist today (Flanagan & Roseman, 2011). This item showed a highly significant overall difference in test values post instruction (χ^2 = 24.00, P = 6.14x10^{-6}, Friedman test; Table 4.16). There was a significant conversion to the correct response in both the post (P = 7.9x10^{-6}, post hoc Friedman Nemenyi test) and retention (P = 0.02, post hoc Friedman Nemenyi test) tests, the correct response (C) remaining as the highest scoring answer. The increase in the correct response was obtained by conversion from all of the other options. There was a degree of waning away from the correct response between the post and retention tests, however, this was not a significant (P = 0.11, post hoc Friedman Nemenyi test). Interestingly the frequency of the attractive distractor A was very low in the pre-test (<5%) and was reduced further post instruction, remaining at an extremely low level (<3%) suggesting this particular alternative conception was not prevalent and was corrected by instruction. Unintentionally response B which stated that most species alive today have existed since life began but with a few species appearing more recently, also seemed to act as an attractive distractor with 18.27% in the pre-test. The frequency of this option decreased in the post-test but then returned to just below pre-test levels suggesting that this alternative conception was not fully accommodated by the students. Finally, the continued decrease in frequency of response D, that there was no way of finding out which species existed since time began, suggests that instruction was successful in helping students appreciate the evidence for evolution.

Table 4.16 Analysis of assessment item 14. Data shown as mean percentages.

Item 14 Response	Pre-test %	Post-test %	Δ post - pre %	Retention %	Δ ret-pre %	Δ post-ret %
A	4.39	2.61	-1.78	2.61	-1.78	0.00
B	18.27	11.27	-7.00	17.08	-1.19	-5.81
C	49.82	74.50	24.67****	66.67	16.84*	7.83
D	25.98	11.39	-14.59	13.17	-12.81	-1.78
U	0.47	0.24	-0.24	0.24	-0.24	0.00
NA	1.07	0.00	-1.07	0.24	-0.83	-0.24

*Notes: green highlighted row showing correct response, orange row showing attractive distractor, * = P ≤ 0.05, **** = P ≤ 0.0001 using Freidman test: post hoc Friedman Nemenyi test.*

Assessment item 11 explored student understanding concerning the extinction of species, the correct response D stating that many species have become extinct throughout the history of life on earth. The other options were all attractive distractors based on alternative conceptions originating from the research of Flanagan and Roseman (2011). Response A representing the alternative conception that very few species have ever become extinct, most are still alive today. Response B representing the alternative conception that except for periodic mass extinction events, extinction is very rare and response C, humans have caused the majority of extinctions, up until recently extinction was rare. Results from the pre-test show that the correct response D (46.62%) was the most frequent followed by response C (27.40%) with the other two distractors at a lower frequency. This item showed a significant overall difference in test values post instruction (F = 3.25, P = 0.04, repeated measures one-way ANOVA; Table 4.17). There was a significant conversion to the correct response in the post-test (P = 0.048) but not from the pre-test to the retention test (P = 0.11) indicating significant waning of understanding after the post test, however, the correct response (D) remained as the highest scoring answer post instruction. The increase in the correct response in the post-test was obtained by conversion from options A and C but not option B which actually increased by 7.8%. These results suggest that instruction was successful in improving the correct understanding of this concept by correcting the alternative conceptions represented by options A and C but that the correct conception was not fully accommodated by the students who returned to their preconceptions. Instruction also had the unintentional effect of increasing the frequency of the alternative conception represented by option B, suggesting that something in teaching interventions reinforced the alternative conception that extinctions only occur in rare mass extinction events rather than throughout the history of the earth.

Table 4.17 Analysis of assessment item 11. Data shown as mean percentages.

Item 11 Response	Pre-test %	Post-test %	Δ post - pre %	Retention %	Δ ret-pre %	Δ post-ret %
A	8.30	4.63	-3.68	6.52	-1.78	-1.90
B	15.30	23.13	7.83	20.76	5.46	2.37
C	27.40	17.91	-9.49	19.45	-7.95	-1.54
D	46.62	53.97	7.35*	52.08	5.46	1.90
U	1.42	0.24	-1.19	0.71	-0.71	-0.47
NA	0.95	0.12	-0.83	0.47	-0.47	-0.36

*Notes: green highlighted row showing correct response, orange rows showing attractive distractors, * = P $\leq$ 0.05*

Finally, to investigate student understanding of how environmental conditions on earth have changed since life began the results from assessment item 15 can be examined. The correct response was D, that conditions have changed dramatically, with some of these changes happening suddenly and others more gradually. Again the other options were attractive distractors based on alternative conceptions originating from the research of Flanagan and Roseman (2011). Response A stating that conditions have stayed the same except for minor fluctuations from year to year. Response B, conditions have remained the same in the oceans but have changed on land and response C, conditions have stayed the same except for a few sudden changes in certain places, such as a meteorite hitting the earth. Pre-instruction the correct option D was the most frequent response (60.38%) with the other options at a much lower level. This item showed a significant overall difference in test values post instruction (χ^2 = 10.47, P = 0.005, Friedman test; Table 4.18), there being a significant conversion to the correct response from the pre-test to the post (P = 0.005, post hoc Friedman Nemenyi test) by conversion from all of the other options, but not from the pre-test to the retention test (P = 0.14, post hoc Friedman Nemenyi test) indicating significant waning of understanding after the post test, however, the correct response (D) remained as the highest scoring answer post instruction. These results suggest that instruction was successful in increasing the correct understanding of this concept by correcting the alternative conceptions represented by option B. However, the alternative conceptions represented by options A and C increased from post-test values in the retention test suggesting that these alternative conceptions were more resistant to change by instruction and that the correct conceptions were not fully accommodated by the students. This was especially apparent in option C which returned to near pre-test value showing the persistence of the alternative conception that conditions have remained the same on earth since life began except for a few sudden changes in certain places.

Table 4.18 Analysis of assessment item 15. Data shown as mean percentages.

Item 15 Response	Pre-test %	Post-test %	Δ post - pre %	Retention %	Δ ret-pre %	Δ post-ret %
A	9.13	5.58	-3.56	7.35	-1.78	-1.78
B	11.27	8.07	-3.20	6.52	-4.74	1.54
C	17.56	14.59	-2.97	17.20	-0.36	-2.61
D	60.38	71.65	11.27**	68.21	7.83	3.44
U	0.12	0.12	0.00	0.24	0.12	-0.12
NA	1.54	0.00	-1.54	0.47	-1.07	-0.47

*Notes: green highlighted row showing correct response, orange rows showing attractive distractors ** = P $\leq$ 0.01 using Freidman test: post hoc Friedman Nemenyi test.*

4.4 Chapter conclusion

The most frequent criticism of the project expressed by teachers during the feedback sessions was that the student questionnaire was too hard and that the scores obtained by their students would not adequately reflect their improved understanding. They worried that their students did not understand the assessment items sufficiently in order to be able to discriminate between the subtle differences in the phrasing of each option. Analysis of data in this chapter suggests that these fears were unfounded on both fronts, as there was a ubiquitous improvement in the percentage of correct responses across the whole assessment instrument and significant changes in the options selected post instruction.

Additionally, if the students showed no understanding of the individual assessment items it would be reasonable to suggest that the students would just randomly guess the answers for each item. However, the responses selected were far from random. Results from the pre-test questionnaires show that the students were able to discriminate between individual assessment item options with a higher proportion of alternative conceptions selected in a homogeneous manner. Post instruction there was an increase the percentage of correct answers across all 15 items in both tests (mean increase of 15.51% ± 8.01 between the pre and post-test, 13.42% ± 5.75 between the pre and retention test) and a decrease in the total frequency of alternative conceptions, suggesting that the teaching intervention programmes provided were successful in correcting student preconceptions across the whole assessment instrument. However, instruction proved relatively more successful in some themes than others; the theme covering the differences between species and common ancestry achieving the largest and most significant improvement post instruction. Conversely students demonstrated a much higher pre-test understanding of the theme covering fossils, extinction and geological time scales but did not make as much improvement post instruction.

Table 4.19 summary of pre-test correct answer percentage and the increases in percentage of correct responses after instruction for all items subdivided by theme.

Assessment item	Pre-test %	Post-pretest %	Ret-pretest %	Theme
13	31.44	10.91****	15.30****	Differences between species and common ancestry
2	69.51	13.64****	14.95****	
3	26.69	28.59****	19.57****	
1	4.39	27.05****	23.01****	
Theme mean	33.01 ± 27.04	20.05 ± 9.07	18.21 ± 3.83	
8	46.50	11.39**	11.86**	Natural selection
4	24.91	18.86****	13.40**	
5	36.42	22.66****	19.81****	
6	23.13	5.69	6.29	

9	18.86	23.01****	18.51****	
7	34.40	12.69**	12.1**	
Theme mean	30.70 ± 10.27	15.72 ± 6.92	13.66 ± 4.93	
10	32.50	11.27**	12.93**	Fossils, extinction and geological time
12	69.87	3.56	3.44	
14	49.82	24.67****	16.84*	
11	46.62	7.35*	5.46	
15	60.38	11.27**	7.83	
Theme mean	51.84 ± 14.17	11.62 ± 7.96	9.30 ± 4.92	

*Note: * = $P \leq 0.05$, ** = $P \leq 0.01$, **** = $P \leq 0.0001$*

The majority of assessment items (1, 2, 3, 4, 5, 7, 8, 9, 10,13 and 14) showed a significant conversion to the correct response from pre-test scores in both the post and retention tests, with no significant difference between the post and retention tests. This suggests that the instruction provided was successful in correcting student preconceptions, this correct understanding being assimilated and retained by the students with no waning in understanding. Two assessment items related to geological time, item 11 (extinctions have occurred throughout geological time not just during mass extinction events) and item 15 (environmental conditions have changed dramatically, with some changes happening suddenly and others more gradually) showed a significant conversion to the correct response from the pre-test to the post-test but no significant difference between the pre and retention test or the post and the retention tests showing that instruction temporarily corrected these alternative conceptions but the students then partially reverted back to their previous preconceptions. This return suggests that the alternative conceptions represented by these assessment items were more resistant to change, the students failing to assimilate the correct conceptions by the instruction provided. Assessment item 12 on the uses of fossils was the only item not to give a significant difference between test scores suggesting the instruction provided did not significantly improve the understanding of this concept as it was already well understood pre-instruction. Individual assessment items grouped according to topic and detail how student conceptions altered after instruction will now be explored.

Although there was a conversion to the scientifically accepted conception in each assessment item, some preconceptions were more resistant to modification than others:

1. Instruction proved very successful in correcting essentialist thinking (in items 13, 8 and 4), moving students away from the concept of species 'essence' and enabling them to appreciate the importance of intraspecific variation in order to understand the process of natural selection.

2. Instruction also proved successful in improving the alternative conception, that members of different species can only share a common ancestor if they share lots of similarities (items 2, 3 and 1). However, students still found this a very challenging part of the assessment, the human and chimpanzee pairing in item 1 proving to be a deeply engrained alternative conception.

3. Anthropomorphic alternative conceptions in which organisms intentionally develop new traits for some specified reason remained persistent and widespread (items 4 and 5) and were particularly resistant to change by instruction,

4. The alternative conception of soft inheritance involving the inheritance of acquired traits (items 6 and 9) also persisted and proved to be very resistant to change by instruction.

5. Lamarckian thinking involving use or disuse influencing inheritance also persisted after instruction in around a quarter of the students (item 5).

6. The teaching intervention programmes seemed to be responsible for an increase in the alternative conception that mass extinction events were responsible for the majority of extinctions (item 11) which was not the intended outcome and probably due to the emphasis of these events in the toilet roll time line.

In the next chapter the importance of teacher and school level effects will be considered enabling the identification of possible confounding or predictive factors.

Chapter 5 Results of teacher and school level analyses comparing both tranches of quantitative data

5.1 Chapter overview

In this chapter the importance of teacher and school level effects are considered to identify possible confounding or predictive factors in both tranches of data. From the analysis of the teacher questionnaires the relationship between their understanding of natural selection and acceptance of the evolutionary process is examined whilst identifying possible sources of ‘taught and learnt’ alternative conceptions.

5.2 Teachers had high acceptance of evolution but demonstrated poor understanding of the concept

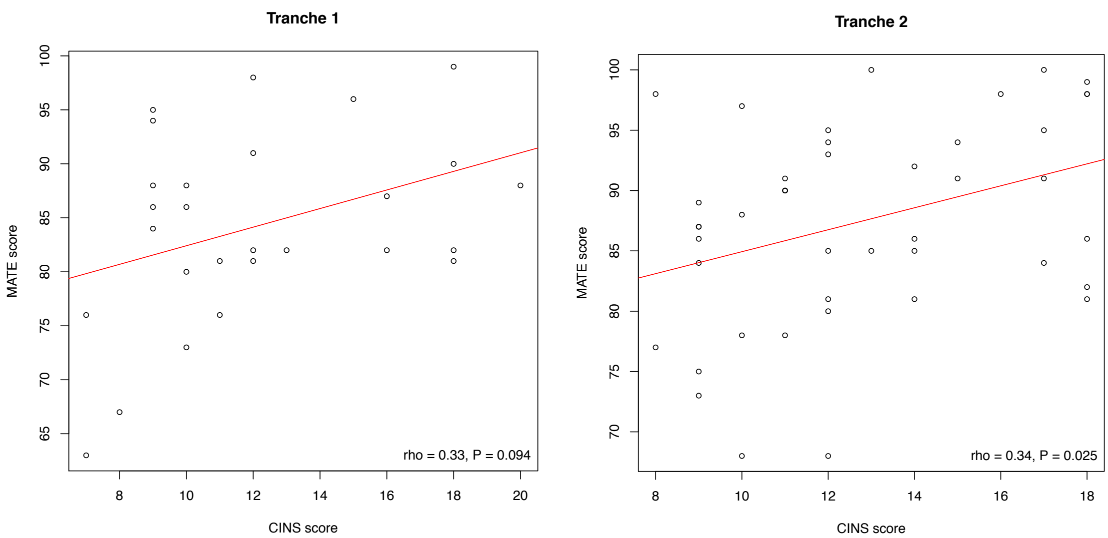
In the previous chapter it was established that students participating in this study held a wide variety of preconceptions some of which were resistant to change by instruction. Multiple studies have shown that many teachers carry out instruction on evolution whilst holding their own related alternative conceptions (Chinsamy & Plagányi, 2008; Nehm & Schonfeld, 2007; Yates & Marek, 2013, 2014) many of which are similar to those held by their students (Bishop & Anderson, 1990; Brumby, 1984; Demastes et al., 1995). Some of these alternative conceptions are retained even after intensive teacher training (Summers & Kruger, 1994).

Whilst it is recognised that fewer primary school teachers are comfortable with teaching evolution (Fowler & Meisels, 2010), there are very few studies focusing on this group of teachers (Glaze & Goldston, 2015). Multiple researchers have called for empirical work to be conducted to explore their understanding and acceptance of evolutionary theory to add to the limited number of studies in this area (Asghar et al., 2007; Nadelson, 2009; Nadelson & Nadelson, 2010; van Dijk & Kattmann, 2009). As the acceptance and knowledge of evolution are sufficiently different constructs they should both be examined when considering an individual’s perspective on evolution (Shtulman, 2006; Southerland, Sinatra, et al., 2001).

The questionnaires completed by participating teachers in this study provides a little more data to this area of research. Completion of the teacher questionnaires was voluntary resulting in 66 out of a possible 76 teachers participating over the two tranches (some teachers present in both tranches).

A positive correlation between the acceptance of evolutionary theory as assessed by the MATE instrument (Rutledge & Warden, 1999) and the understanding of natural selection assessed by using the CINS instrument (Anderson et al., 2002) was shown in both tranches (Fig. 5.1). Both correlations showed very similar moderate effect sizes but was only statistically significant in the larger tranche of data (ρ = 0.33, P = 0.094, Spearman's rank correlation, tranche 1; ρ = 0.34, P = 0.025, tranche 2, Spearman's rank correlation). However, a significant combined P result (P = 0.016, Fisher's test) was obtained suggesting a significant positive correlation between the acceptance of evolution and the understanding of natural selection in this study's cohort of teachers.

Figure 5.1 Scatterplots with lines of regression showing correlation between MATE and CINS scores for participating teachers.



Notes: Tranche 1 n = 27; Tranche 2 n = 44

Teachers in both tranches achieved mean MATE scores reflecting a high acceptance (Rutledge, 1996) of evolution (84.30 ± 8.71 tranche 1; 87.45 ± 8.30 tranche 2), there being no significant difference between MATE scores in the two tranches (w = 469, P = 0.14, Wilcoxon rank sum test) with a small effect size (Cliff's delta = 0.21). When the MATE assessment instrument was broken down by concept (Table 6.1) and responses to individual items examined, a fairly even distribution of high acceptance scores across the concepts in both tranches was observed (Tables 5.2 and 5.3).

Table 5.1 Breakdown of the MATE assessment instrument, according to concept covered and question

Concept	Item numbers
Process of evolution	1,9,18, 19
Scientific validity of evolutionary theory	2,10,12,13,14, 20
Evolution in humans	3,15
Evidence for evolution	4,6,8,16
Scientific community's view of evolution	5,17
Age of the Earth	7,11

Table 5.2. Teacher responses to sub-scale items pertaining to the MATE assessment instrument in Tranche 1

Statement	**Percentage response in Tranche 1**					
	Strongly disagree	**Disagree**	**Undecided**	**Agree**	**Strongly agree**	**NA**
1. Organisms existing today are the result of evolutionary processes that have occurred over millions of years	0.00	0.00	0.00	11.11	88.89	0.00
2. The theory of evolution is incapable of being scientifically tested.	70.37	22.22	11.11	11.11	3.70	0.00
3. Modern humans are the product of evolutionary processes that have occurred over millions of years	0.00	3.70	3.70	22.22	70.37	0.00
4. The theory of evolution is based on speculation and not valid scientific observation and testing	44.44	44.44	0.00	7.41	3.70	0.00
5. Most scientists accept evolutionary theory to be a scientifically valid theory.	0.00	7.41	3.70	59.26	29.63	0.00
6. The available data are ambiguous (unclear) as to whether evolution actually occurs	22.22	55.56	14.81	3.70	3.70	0.00
7. The age of the earth is less than 20,000 years	62.96	18.52	11.11	3.70	3.70	0.00
8. There is a significant body of data that supports evolutionary theory	0.00	0.00	0.00	55.56	40.74	3.70
9. Organisms exist today in essentially the same form in which they always have	29.63	51.85	3.70	7.41	3.70	3.70
10. Evolution is not a scientifically valid theory	48.15	40.74	11.11	0.00	0.00	0.00
11. The age of the earth is at	0.00	0.00	18.52	48.15	33.33	0.00

least 4 billion years						
12. Current evolutionary theory is the result of sound scientific research and methodology	0.00	0.00	0.00	66.67	33.33	0.00
13. Evolutionary theory generates testable predictions with respect to the characteristics of life	0.00	3.70	14.81	70.37	11.11	0.00
14. The theory of evolution cannot be correct since it disagrees with the Biblical account of creation	62.96	29.63	3.70	3.70	0.00	0.00
15. Humans exist today in essentially the same form in which they always have.	40.70	40.70	11.11	7.41	0.00	0.00
16. Evolutionary theory is supported by factual historical and laboratory data	0.00	3.70	11.11	59.26	25.93	0.00
17. Much of the scientific community doubts if evolution occurs.	37.04	55.56	3.70	0.00	0.00	3.70
18. The theory of evolution brings meaning to the diverse characteristics and behaviours observed in living forms.	0.00	0.00	3.70	51.85	44.44	0.00
19. With few exceptions, organisms on earth came into existence at about the same time	40.74	33.33	11.11	14.81	0.00	0.00
20. Evolution is a scientifically valid theory	0.00	3.70	3.70	44.44	43.15	0.00

Notes: Colour coding according to concept (see Table 6.3). Mean percentage values shown.

Table 5.3. Teacher responses to sub-scale items pertaining to the MATE assessment instrument in Tranche 2.

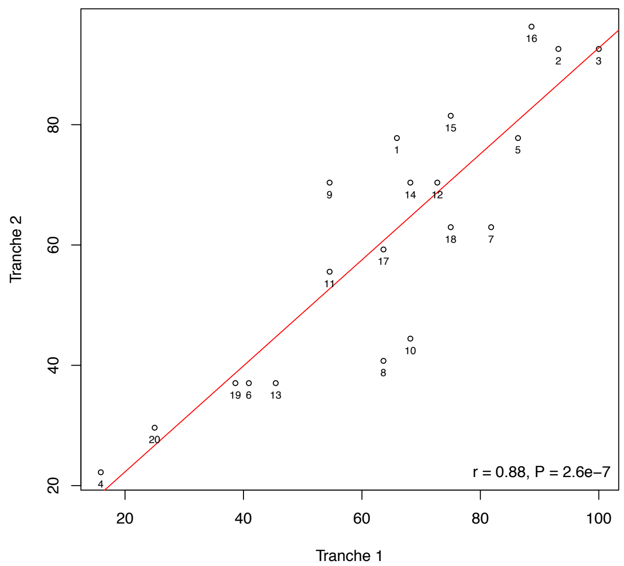
Statement	**Percentage response in Tranche 2**					
	Strongly disagree	**Disagree**	**Undecided**	**Agree**	**Strongly agree**	**NA**
1. Organisms existing today are the result of evolutionary processes that have occurred over millions of years	0.00	0.00	0.00	11.36	88.64	0.00
2. The theory of evolution is incapable of being scientifically tested.	88.64	11.36	0.00	0.00	0.00	0.00
3. Modern humans are the product of evolutionary processes that have occurred over millions of years	2.27	2.27	0.00	18.18	77.27	0.00
4. The theory of evolution is based on speculation and not valid scientific observation and testing	50.00	38.64	4.55	6.82	0.00	0.00

5. Most scientists accept evolutionary theory to be a scientifically valid theory.	0.00	0.00	9.09	40.91	50.00	0.00
6. The available data are ambiguous (unclear) as to whether evolution actually occurs	27.27	52.27	15.91	4.55	0.00	0.00
7. The age of the earth is less than 20,000 years	79.55	9.09	9.09	2.27	0.00	0.00
8. There is a significant body of data that supports evolutionary theory	0.00	4.55.27	2.27	34.09	59.09	0.00
9. Organisms exist today in essentially the same form in which they always have	34.09	45.45	6.82	6.82	6.82	0.00
10. Evolution is not a scientifically valid theory	61.36	29.55	2.27	4.55	2.27	0.00
11. The age of the earth is at least 4 billion years	0.00	2.27	9.09	36.36	52.27	0.00
12. Current evolutionary theory is the result of sound scientific research and methodology	0.00	2.27	4.55	50.00	43.18	0.00
13. Evolutionary theory generates testable predictions with respect to the characteristics of life	0.00	0.00	11.36	56.82	31.82	0.00
14. The theory of evolution cannot be correct since it disagrees with the Biblical account of creation	79.55	18.18	0.00	2.27	0.00	0.00
15. Humans exist today in essentially the same form in which they always have.	36.368	50.00	4.55	6.28	2.27	0.00
16. Evolutionary theory is supported by factual historical and laboratory data	0.00	2.27	4.55	54.55	38.64	0.00
17. Much of the scientific community doubts if evolution occurs.	47.73	40.91	9.09	2.27	0.00	0.00
18. The theory of evolution brings meaning to the diverse characteristics and behaviours observed in living forms.	0.00	0.00	2.27	59.09	38.64	0.00
19. With few exceptions, organisms on earth came into existence at about the same time	38.64	59.09	2.27	0.00	0.00	0.00
20. Evolution is a scientifically valid theory	0.00	0.00	0.00	31.82	68.18	0.00

Notes: Colour coding according to concept (see Table 6.3). Mean percentage values shown.

There was also no significant difference (w = 522.5, P = 0.40, Wilcoxon rank sum test) with a negligible effect size (Cliff's delta = 0.12) between the CINS scores for both tranches with very similar means (12.91 ± 3.81 tranche 1; 12.77 ± 3.26 tranche 2). However, the scores for both tranches fell short of the cut off value suggested by Anderson et al. (2010) of 16/20 or higher to represent a sound understanding of natural selection. A correlation was carried out on the mean percentage of correct responses for each of the 20 assessment items to examine the relationship between the two tranches of data. A statistically significant positive correlation, with a large effect size was found between the two tranches of data ($r = 0.88$, $P = 2.61 \times 10^{-7}$, Pearson's product moment correlation; Fig.5.2) indicating that the responses obtained in both tranches of data were consistent showing reproducibility.

Figure 5.2 Scatterplot with line of regression showing correlation between mean percentage of CINS scores for each of the 20 assessment items in the two tranches of data.



The extent and nature of the alternative conceptions held by teachers in this project can be explored further by comparing the mean percentage of each response option within each assessment item. A nominal cut off value of < 50% correct responses was used to identify assessment items which represented areas of poor understanding by the teachers. From the twenty assessment items only five met this criterion (items 4, 6, 13, 19 and 20) and it is these items that we will examine in more depth. Surprisingly, all of these selected assessment items fell into a limited number of related scientific concepts; change in a population, origin of species and the origin of variation.

Assessment items 4 and 13 investigated the understanding of how changes in a population occur. The scientifically correct conception is that due to intraspecific variation there is differential survival and reproduction of individuals within a population, leading to a gradual change in the population, with the proportion of individuals with favourable traits accumulating over the generations (options 4B and 13B; Tables 5.4 and 5.5). However, a large proportion of teachers in both tranches selected the common alternative conception (option D in both items) that mutations occurred in the populations to meet the needs of the population, an example of external anthropomorphism in which the environment is acting as the intentional agent. Additionally, the selection of option C in both tranches reflected the choice of explanations involving the inheritance of acquired traits or soft inheritance.

Table 5.4. Teacher responses to sub-scale items pertaining to the CINS assessment item 4 in both tranches of data.

Assessment item 4. In the finch population, what are the primary changes that occur gradually over time?		Percent response	
		Tranche 1	Tranche 2
A	The traits of each finch within a population gradually change.	7.41	13.64
B	The proportions of finches having different traits within a population change.	22.22	15.91
C	Successful behaviours learned by finches are passed on to offspring.	22.22	27.27
D	Mutations occur to meet the needs of the finches as the environment changes	48.15	40.91

Table 5.5. Teacher responses to sub-scale items pertaining to the CINS assessment item 13 in both tranches of data

Assessment item 13. In guppy populations, what are the primary changes that gradually occur over time?		Percent response	
		Tranche 1	Tranche 2
A	The traits of each individual guppy within a population gradually change.	7.41	9.09
B	The proportions of guppies having different traits within a population change.	37.04	45.45
C	Successful behaviours learned by certain guppies are passed on to offspring.	18.52	11.36
D	Mutations occur to meet the needs of the guppies as the environment changes	37.04	34.09

Notes: Green rows showing correct response and orange rows showing attractive distractors. Mean percentages shown.

Assessment items 6 and 19 investigated the understanding of the origin of variation, the correct conception being that random mutations and sexual reproduction produce variations; while many are harmful or neutral, a few are beneficial in some environments (options 6B and 19C; Tables 656 and 5.7). A large proportion of the teachers selected distractors that represented internal anthropomorphic thinking (options 6A and 19A), that the change was intentional and caused by the organisms need to change genetically. Another common alternative conception in this pair of items was also related to intentionality however, in this instance the adaptive responses were due to specific environmental agents (options 6C and 19D).

Table 5.6. Teacher responses to sub-scale items pertaining to the CINS assessment item 6 in both tranches of data

Assessment item 6. How did the different beak types first arise in the Galapagos finches?		Percent response	
		Tranche 1	Tranche 2
A	The changes in the finches' beak size and shape occurred because of their need to be able to eat different kinds of food.	44.44	29.55
B	Changes in the finches' beaks occurred by chance, and when there was a good match between beak structure and the available food, those birds had more offspring.	37.04	40.91
C	The changes in the finches' beaks occurred because the environment induced the desired genetic features.	11.11	11.36
D	The finches' beaks changed a little bit in shape and size with each successive generation, some getting larger and some getting smaller.	7.41	18.18

Table 5.7. Teacher responses to sub-scale items pertaining to the CINS assessment item 19 in both tranches of data

Assessment item 19. According to the theory of natural selection, where did the variation in body size in the three species of lizards most likely come from?		Percent response	
		Tranche 1	Tranche 2
A	The lizards needed to change in order to survive, so beneficial new traits developed.	29.63	29.55
B	The lizards wanted to become different in size, so beneficial new traits gradually appeared in the population.	3.70	4.55
C	Random genetic and sexual recombination both created new variations.	37.04	38.64
D	The island environment caused the genetic changes in the lizards	25.93	20.45

Notes: Green rows showing correct response and orange rows showing attractive distractors. Mean percentages shown.

And finally, assessment item 20 investigated the understanding of the origin of species; that an isolated population may change so much over time that it becomes a new species (option B; Table 5.8). In both tranches option D was the most frequent selection which together with option A representing another example of internal anthropomorphic thinking in which the organism intentionally becomes a new species over time because it needs to. Interestingly, there was a very low frequency of option C representing essentialist thinking in our teachers which was confirmed by the results of item 8 (option B).

Table 5.8. Teacher responses to sub-scale items pertaining to the CINS assessment item 20 in both tranches of data.

Assessment item 20. What would cause one species to change into three species over time?		Percent response	
		Tranche 1	Tranche 2
A	Groups of lizards encountered different island environments so the lizards needed to become new species with different traits in order to survive.	7.41	9.09
B	Groups of lizards must have been geographically isolated from other groups and random genetic changes must have accumulated in these lizard populations over time.	29.63	25.00
C	There may be minor variations, but all lizards are essentially alike and all are members of a single species.	0.00	2.27
D	In order to survive, different groups of lizards needed to adapt to the different islands, and so all organisms in each group gradually evolved to become a new lizard species.	59.26	56.82

Notes: Green row showing correct response and orange rows showing attractive distractors. Mean percentages shown.

In summary, these results suggest that whilst the teachers in this study showed a high acceptance of evolution they did not have an adequate understanding of the process of natural selection. This lack of understanding mainly resulting from widely held alternative conceptions relating to anthropomorphic thinking; that individual organisms or populations change intentionally because they need to, these changes either directed by the organism itself (internal) or by the environment (external). There was also evidence that some teachers also held alternative conceptions related to soft inheritance, in that new traits acquired during an organism's life time could be passed onto future generations. These findings mirror those reported in their students post instruction in the previous chapter and demonstrates the persistence of these deeply engrained alternative conceptions into adulthood. This also suggests that the training provided pre-instruction may not have adequately addressed these alternative conceptions in the teachers which could imply that they were unable to correct these in their classes or even may have passed these incorrect conceptions on to their students. However, this is purely speculative, as the teachers only completed one full questionnaire before teaching their classes and so it is conceivable that their understanding of the topic may have been altered by actually using the resources and teaching their classes (Schank, 1995).

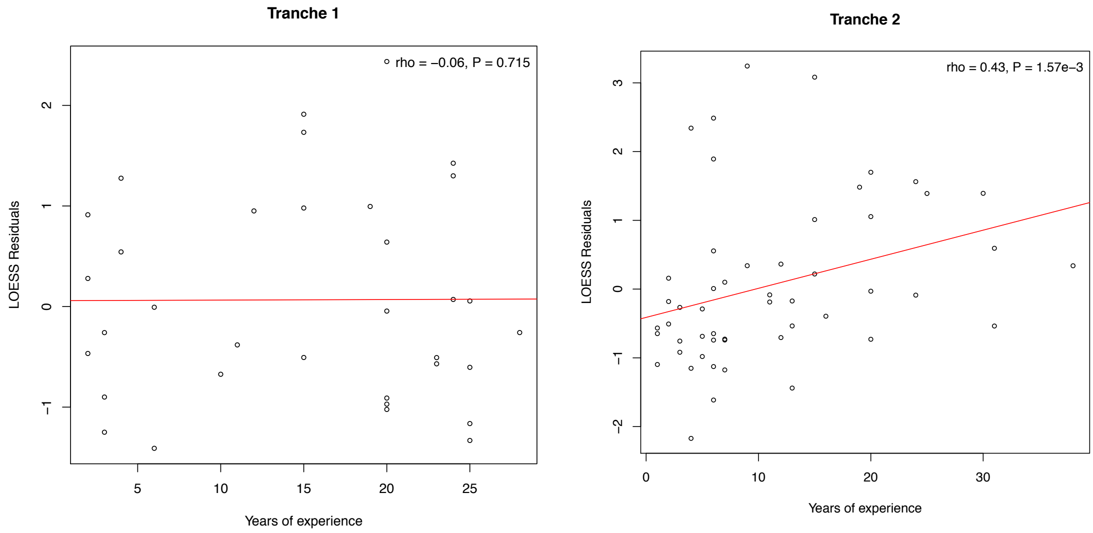
5.3 Class level analyses

All previous analyses performed (at student level) come with the caveat that there may be pseudoreplication of data as students within any given class share the same teacher, thus introducing a component of non-independence between the students. For this reason, and to consider the importance of teacher level effects, class level analyses were carried out using mean class difference scores that were matched to completed teacher questionnaires ($n = 34$ in tranche 1, $n = 50$ in tranche 2). Classes without a corresponding completed teacher questionnaire were omitted from this level of analyses. As these analyses were based on the changes in scores between tests, LOESS residuals were employed to correct for pre-test and mitigate for ceiling effect rather than using uncorrected raw scores. These mean class LOESS residual scores were then used as a measure of teacher effectiveness in the classroom. Additionally, the tranches of data were treated separately to allow comparison between them. These class level analyses enabled us to study a series of possible confounding or predicting factors previously identified in the research literature (see chapter 1).

5.3.1 Years of teaching experience was a predictor of teacher effectiveness

The results for this analysis, the correlation between mean class LOESS residuals and years of teaching experience were quite different between the 2 tranches of data. There was a slight negative correlation which was not significant with a negligible effect size in tranche 1 ($\rho = -0.06$, $P = 0.715$, Spearman's rank correlation) whist there was a statistically significant positive correlation with a medium effect size in tranche 2 ($\rho = 0.43$, $P = 1.57 \times 10^{-3}$, Spearman's rank correlation); see Fig.5.3 for comparison. From the literature we would expect to find a curvilinear trend in which less experienced teachers are less effective than more experienced ones and then the benefits of experience levelling off after about 5 years (Rosenholtz, 1986) due to lack of Continued Professional Development training, disillusionment or tiredness in older teachers. Our results show that there was no significant difference in effectiveness between less experienced and more experienced teachers in tranche 1, suggesting that all teachers in this cohort were equally well prepared to teach their students regardless of their years of experience (Denton & Peters, 1988). However, in the second tranche there seemed to be a continued improvement in effectiveness as the teachers became more experienced suggesting that there was no levelling off of performance in the veteran teachers. Overall, a significant combined P value of 9.00×10^{-3} (Fisher's test) was obtained suggesting that years of experience was a predictor of student performance with more experienced teachers performing better than their less experienced colleagues.

Figure 5.3 Scatterplots with lines of regression showing correlation between mean class LOESS residual scores and years of teaching experience in both tranches.

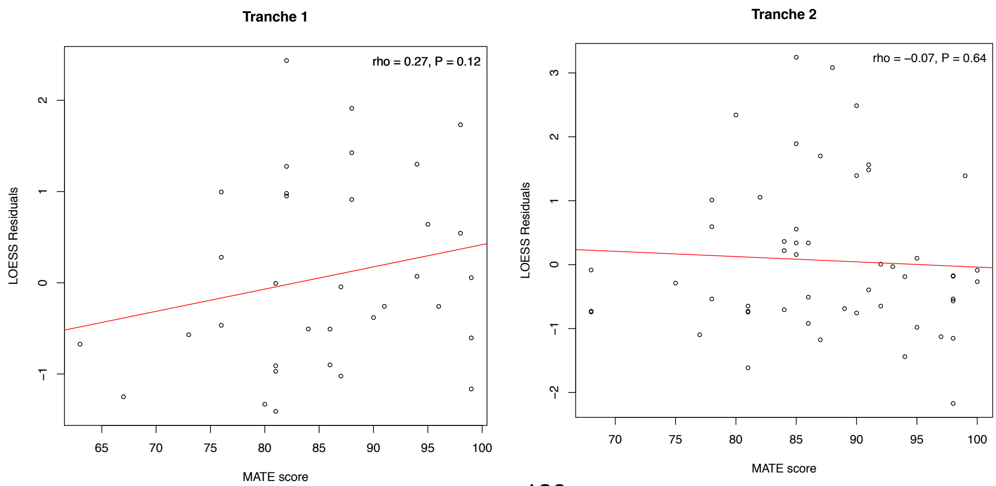


Notes: Tranche 1, n = 34, mean years of experience = 15.09 ± 8.70. Tranche 2, n = 50, mean years of experience = 11.29 ± 9.03.

5.3.2 Acceptance of evolution was not a predictor of effectiveness

When the mean class LOESS residual scores and MATE scores were correlated a positive correlation was found in tranche 1 (ρ = 0.27, P = 0.12, Spearman's rank correlation) but a negative correlation in tranche 2 (ρ = -0.07, P = 0.64, Spearman's rank correlation). However, both had small effect sizes with a non-significant combined P value (P = 0.28, Fisher's test; Fig.5.4). This is rather surprising as teachers with higher acceptance levels of evolution would be expected to be more effective at teaching this topic, maybe approaching the topic in a more open and enthusiastic manner. The results of this analysis suggest that acceptance of evolution was not a predictor of their effectiveness to teach the topic in class in the teachers participating in this study.

Figure 5.4 Scatterplots with lines of regression showing correlation between mean class LOESS residual scores and MATE scores in both tranches.

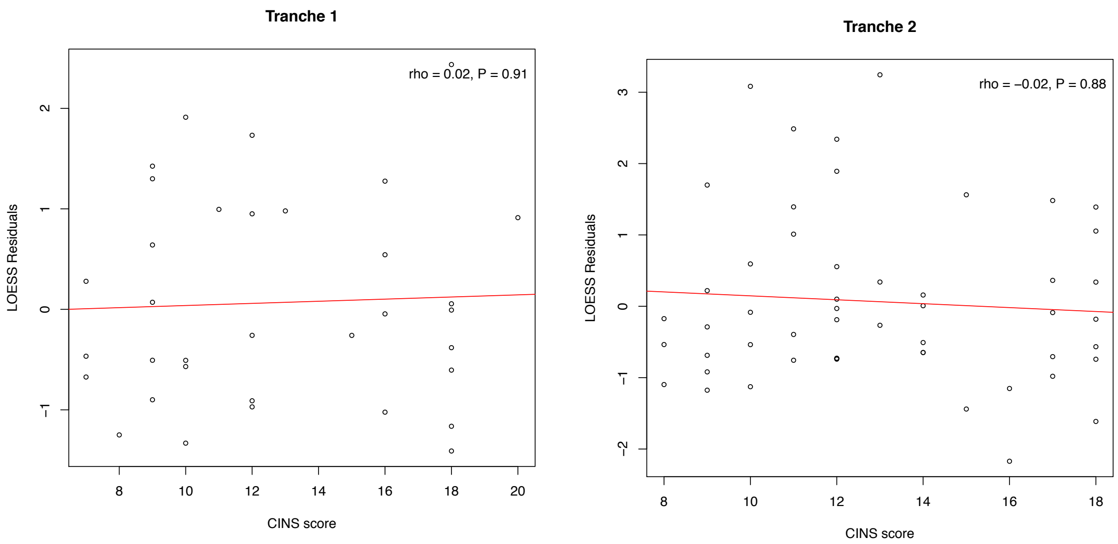


Notes: Tranche 1, n = 34, mean MATE score = 85.59 ± 9.07. Tranche 2, n = 50, mean MATE score = 87.38 ± 8.32.

5.3.3 Understanding of natural selection was not a predictor of effectiveness

Do teachers with a better understanding of natural selection make for more effective teachers? In order to test this question, the mean class LOESS residual scores were correlated with their CINS scores. A positive correlation in tranche 1 (ρ = 0.02, P = 0.91, Spearman's rank correlation) was obtained but a negative correlation in tranche 2 (ρ = -0.02, P = 0.88, Spearman's rank correlation); see Fig. 5.5 for comparison. Both of these tranches had negligible effect sizes with a non-significant combined P value (P = 0.98, Fisher's test) suggesting that there was no difference in effectiveness of the teachers. This result is perhaps surprising as teachers with a better understanding of natural selection might be expected to be more effective as they would be better equipped to answer student questions and clarify potential conceptual conflicts (Driver, 1994).

Figure 5.5 Scatterplots with lines of regression showing correlation between mean class LOESS residual scores and CINS scores in both tranches.

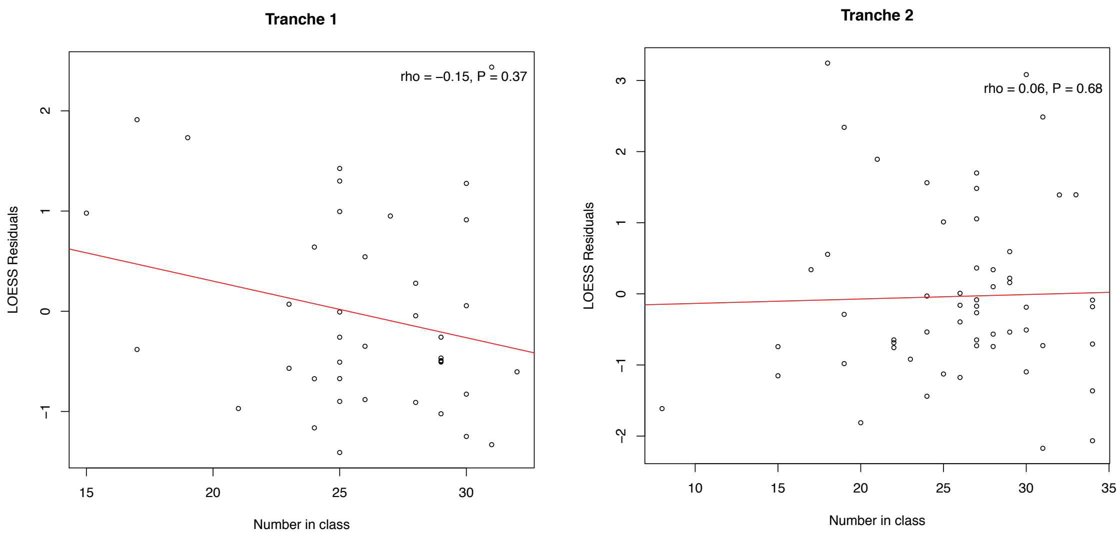


Notes: Tranche 1, n = 34, mean CINS score = 12.71 ± 4.03. Tranche 2, n = 50, mean CINS score = 12.94 ± 3.23.

5.3.4 Number of students in the class did not affect performance

In order to study the effect of class size on teacher performance another correlation was carried out, in which students in smaller classes were expected to perform better than those in larger classes. In this analysis a negative correlation was obtained in tranche 1 (ρ = -0.15, P = 0.37, Spearman's rank correlation) but a positive correlation in tranche 2 (ρ = 0.06, P = 0.68, Spearman's rank correlation). Both tranches had negligible effect sizes with a non-significant combined P value (P = 0.60, Fisher's test; Fig. 5.6). These results suggest that class size was not a predictor of teacher effectiveness for this topic.

Figure 5.6 Scatterplots with lines of regression showing correlation between mean class LOESS residual scores and class size in both tranches.



Notes: Tranche 1, n = 34, mean class size = 25.98 ± 4.02. Tranche 2, n = 50, mean class size = 25.84 ± 5.50.

5.3.4 Difference in teacher evolution confidence scores was a predictor of effectiveness

Three questions on the teacher questionnaires specifically targeted their confidence in teaching specific parts of the national curriculum pertaining to evolution. Each of the responses to these three questions was based on a five-point Likert scale allowing a maximum of fifteen marks.

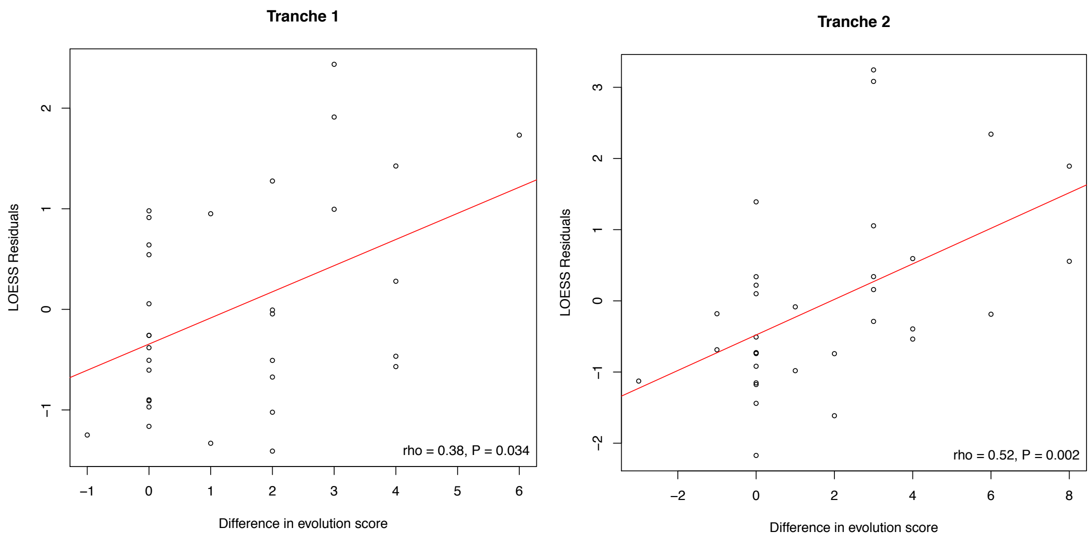
Q1. How confident are you about teaching students how to “Recognise that living things have changed over time and that fossils provide information about living things that inhabited the Earth millions of years ago”.

Q2. How confident are you about teaching students how to “Recognise that living things produce offspring of the same kind, but normally offspring vary and are not identical to their parents”.

Q3. How confident are you about teaching students how to “Identify how animals and plants are adapted to suit their environment in different ways and that adaptation may lead to evolution”.

Scores were obtained from the teachers before and after teaching the topic on separate forms so that we could assess their unbiased perceived change in confidence gained from participation in the study. Some teachers who completed the pre-teaching questionnaires did not complete the matching post-test resulting in a slightly smaller data set in both tranches. The changes in confidence level were then correlated with the mean class LOESS residual scores to see if the two were related; Fig. 5.7)

Figure 5.7 Scatterplots with lines of regression showing correlation between mean class LOESS residual scores and perceived change in confidence in both tranches.



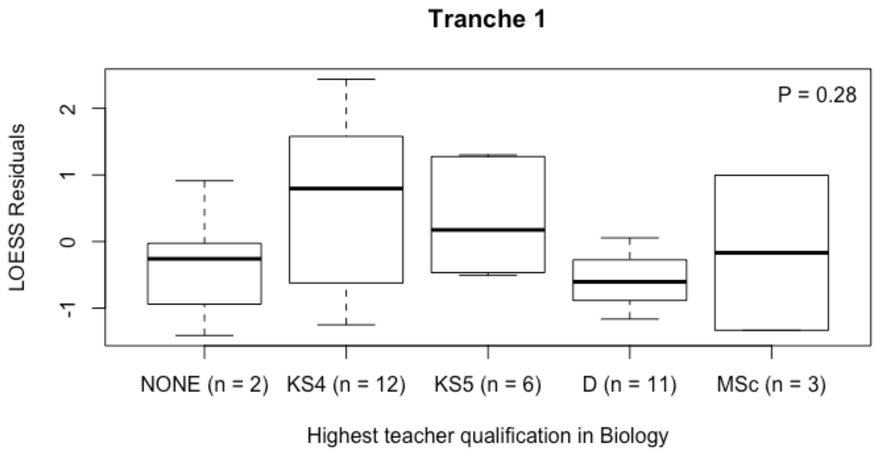
Notes: Tranche 1, n = 32, mean difference in evolution confidence = 1.44 ± 1.72. Tranche 2, n = 33, mean difference in evolution confidence = 1.79 ± 2.61.

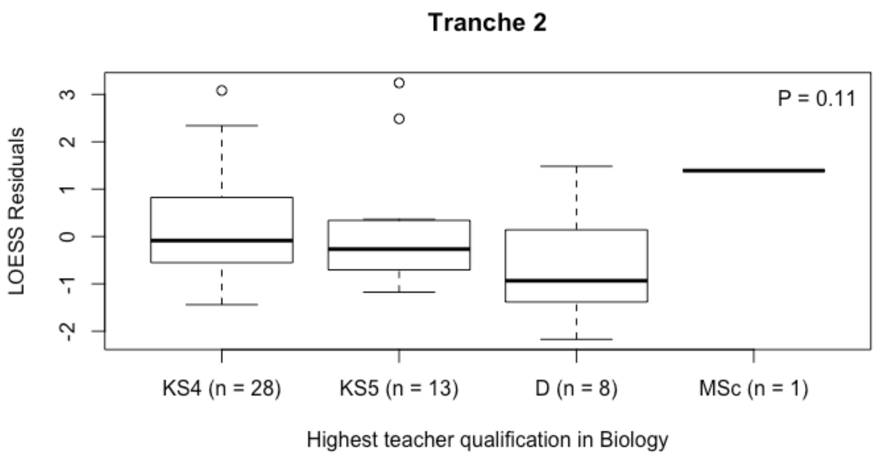
Both tranches of data gave very similar results for this factor, with statistically significant positive correlations between mean class LOESS residual scores and perceived changes in confidence levels (ρ = 0.38, P = 0.03, Spearman’s rank correlation, tranche 1; ρ = 0.52, P = 0.02, Spearman's rank correlation, tranche 2). There being a medium effect size in tranche 1 and a large effect size in tranche 2 with a significant combined P value (P = 0.007, Fisher’s test). This suggests that teachers who felt more confident in their ability to teach the topic were more successful at improving their students’ understanding.

5.3.5 Highest qualifications in Biology and formal lessons in evolution were not predictors of effectiveness

Teachers with a higher level of Biological education would be expected to be more effective at teaching this topic. In both tranches there was no significant difference between the highest Biology qualification in held by the teachers and their mean class LOESS residual score, with a combined P value of 0.14 when a Fisher's test was carried out (χ^2 = 5.08, P = 0.28, Kruskal-Wallis rank sum test, tranche 1; χ^2 = 6.02, P = 0.1, Kruskal-Wallis rank sum test, tranche 2; Fig.5.8). However, both tranches had medium effect sizes (ε^2 = 0.13, tranche 1; ε^2 = 0.11, tranche 2).

Figure 5.8 Boxplots showing relationship between mean class LOESS residual scores and highest qualification in Biology held in both tranches.

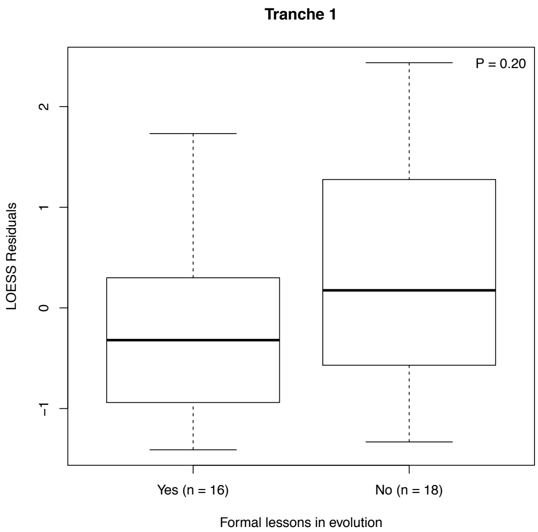


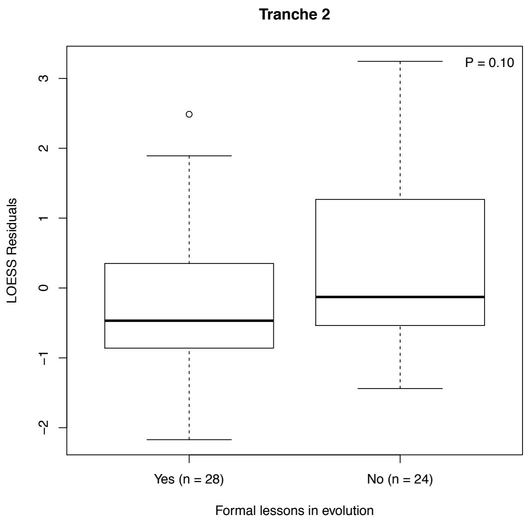


Notes: Qualification levels, KS4 = GCSE/O level, KS5 = A Level, D = degree.

Similarly, teachers who had received formal instruction on this topic would be expected to be more effective in the classroom than teachers with no formal instruction. However, receiving formal instruction on evolution during their own education did not seem to be a predictor of teacher effectiveness with no significant difference between the two groups of teachers in either tranche (w = 106, P = 0.20, Wilcoxon rank sum test, tranche 1; w = 245, P = 0.10, Wilcoxon rank sum test, tranche 2; Fig.5.9) or their combined P value (P = 0.09 Fisher's test). Both tranches also had small effect sizes (Cliff's d = -0.26 tranche 1; Cliff's d = -0.27 tranche 2).

Figure 5.9 Boxplots showing relationship between mean class LOESS residual scores and whether the teachers had experienced any formal instruction on evolution in both tranches.

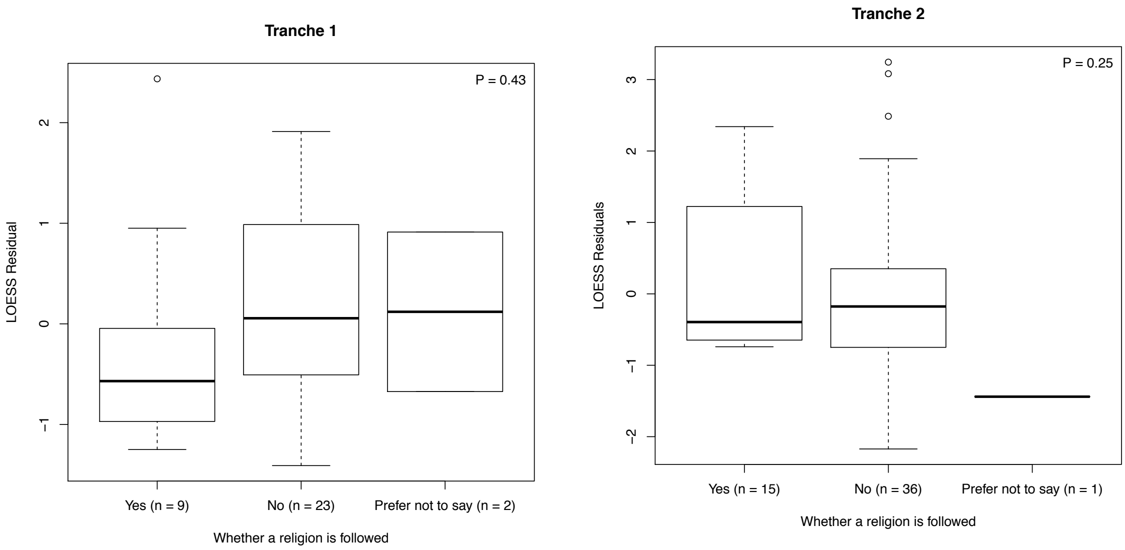




5.3.6. Religiousness was not a barrier to effectiveness nor to the acceptance of evolution

Teachers holding religious beliefs could be less effective at teaching evolution due to their personal belief systems. To test this hypothesis, religiousness was examined to ascertain whether it was a confounding factor. In both tranches the relationship between whether the teacher followed a religion and their mean class LOESS residual score was not significant (χ^2 = 1.68, P = 0.43, Kruskal-Wallis rank sum test, tranche 1; χ^2 = 2.81, P = 0.25, Kruskal-Wallis rank sum test, tranche 2; Fig. 5.10) with a combined P value of 0.34 (Fisher's test) and small effect sizes in both tranches (ε^2 = 0.04, tranche 1; ε^2 = 0.05, tranche 2). Individual teacher's religiousness and MATE scores (as a measure of their acceptance of evolution) were then compered to examine whether this could be contributing to the previous result. No statistically significant relationship between these factors was found (χ^2 = 4.22, P = 0.12, Kruskal-Wallis rank sum test, tranche 1; χ^2 = 3.21, P = 0.20, Kruskal-Wallis rank sum test, tranche 2; combined P = 0.11, Fisher's test), with a medium effect size in tranche 1 and a small one in tranche 2 (ε^2 = 0.11, tranche 1; ε^2 = 0.06, tranche 2). These results suggest that the religious beliefs held by teachers participating in this study did not affect their effectiveness to teach the topic or their acceptance of evolution.

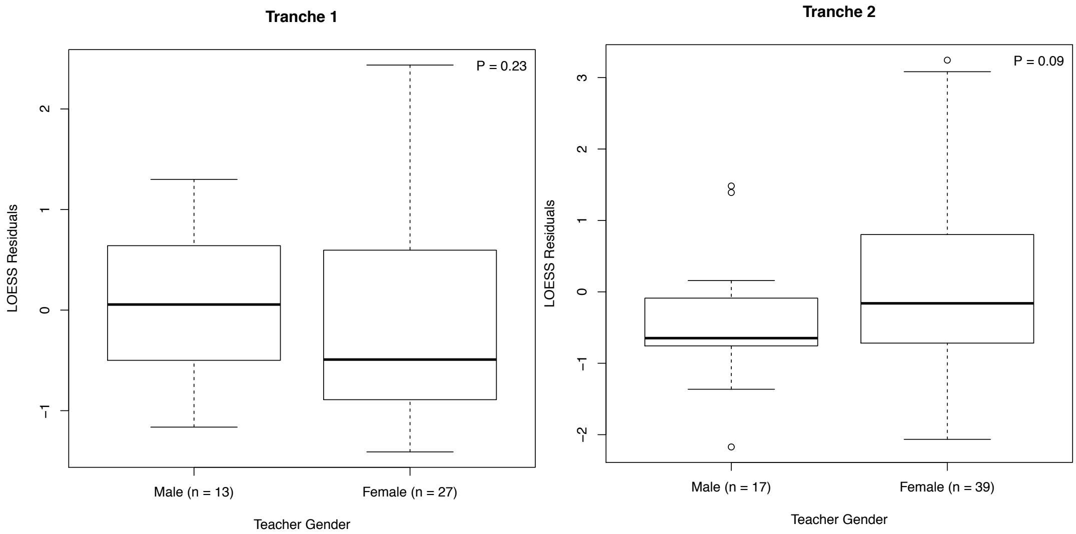
Figure 5.10 Boxplots showing relationship between mean class LOESS residual scores and whether the teacher followed a religion in both tranches.



5.3.7. Teacher gender was not a predictor of effectiveness

An investigation into whether teacher gender could affect student performance in this female biased sample was also carried out (n = 13 males, n = 27 females, tranche 1; n = 17 males, n = 39 females, tranche 2). The analysis found that the relationship between teacher gender and their mean class LOESS residual score was not a significant predictor of effectiveness (w = 218, P = 0.23, Wilcoxon rank sum test, tranche 1; w = 235, P = 0.09, Wilcoxon rank sum test, tranche 2; combined P = 0.09, Fisher's test; Fig. 5.11), with both tranches reporting small effect sizes (Cliff's d = 0.24, tranche 1; Cliff's d = 0.29, tranche 2). These results suggesting that male and female teachers were equally as effective in delivering this topic in the classroom.

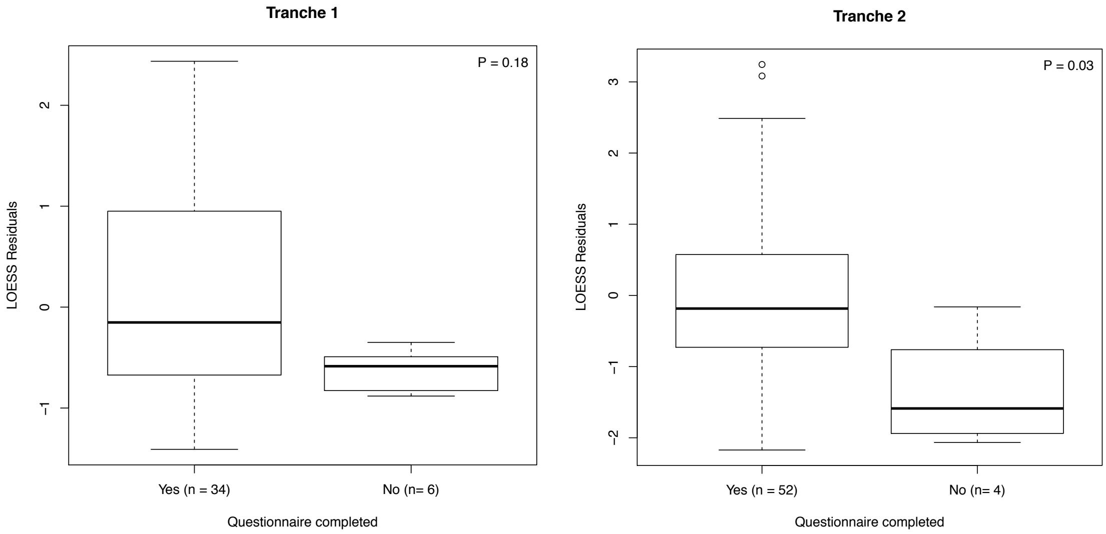
Figure 5.11 Boxplots showing relationship between mean class LOESS residual scores and teacher gender in both tranches



5.3.8. Completion of the teacher questionnaire was a predictor of effectiveness

A further analysis was carried out to check whether teacher effectiveness was related to the completion of the teacher questionnaires. In some of the larger primary and middle schools the 'lead teacher' signed up their whole team for participation. Whilst these 'co-opted' team teachers completed the instruction and carried out the student assessments, a small number of these teachers did not complete their voluntary teacher questionnaires. Their reluctance to complete their own questionnaires could be caused by a lack of confidence in their content related knowledge or lower motivation levels to complete the extra work imposed upon them. In this analysis the relationship between the mean class LOESS residual scores for all classes and whether or not the pre-teaching questionnaire was completed was explored. In tranche 1 there was a non-significant difference between the mean class LOESS residual scores with a medium effect size (w = 138, P = 0.18, Wilcoxon rank sum test, Cliff's d = 0.35) and whether the questionnaire was completed, whilst in tranche 2 a significant difference with a large effect size was found with those not completing the questionnaires being less effective in teaching the topic (w = 175, P = 0.03, Wilcoxon rank sum test, Cliff's d = 0.68); see Fig. 5.11 for comparison. A combined P value of 0.03 (Fisher's test) was obtained confirming evidence of completion bias; teachers who did not complete their questionnaires being less effective in improving student understanding than their colleagues who completed them.

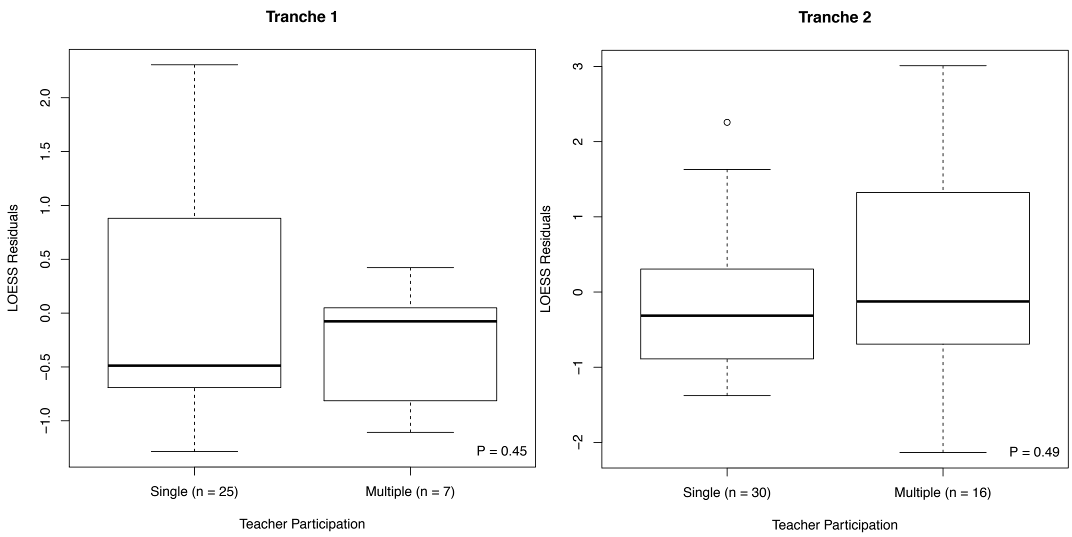
Figure 5.11 Boxplots showing relationship between mean class LOESS residual scores and whether teacher questionnaires were completed in both tranches.



5.3.9. Multiple participation was not a predictor of effectiveness

One final analysis was carried out to examine whether individual teachers who taught multiple classes within the same year or over different years were more effective than other teachers who only taught the topic once as proposed by the research of Andersson and Bach (2005). It would be expected that the more familiar a teacher is with the topic and resources provided, the greater their improvement in mean class LOESS residual scores. However, there was no significant difference between these two groups of teachers in either tranche (w = 105, P = 0.45, Wilcoxon rank sum test, tranche 1; w = 209, P = 0.49, Wilcoxon rank sum test, tranche 2) with a combined P value of 0.55 (Fisher's test); Fig.5.12. This result suggests that multiple participation in this study was not a predictor of teacher effectiveness.

Figure 5.12 Boxplots showing relationship between mean class LOESS residual scores and number of times the teacher participated in both tranches.



5.3.10 All significant predictors in the teacher level analysis failed multi-test correction.

In the teacher level analyses only three significant factors were identified with combined P values of ≤ 0.05, these being a teacher's years of experience, their perceived improvement in confidence to teach the topic and whether or not they completed their pre-teaching questionnaire. It is interesting to note that all three factors are possibly related to teacher confidence either directly or indirectly. However, when multi-test corrections are carried out, these three significant predictors fail to pass either the Holm or Bonferroni correction tests; Table 5.9.

Table 5.9. Table of comparison for teacher level multi-test adjusted significance levels.

Factor	Combined P value obtained	Bonferroni adjusted P value	Holm adjusted P value
Increase in confidence	0.007	0.077	0.077
Years of experience	0.009	0.099	0.090
Completion bias	0.030	0.330	0.270

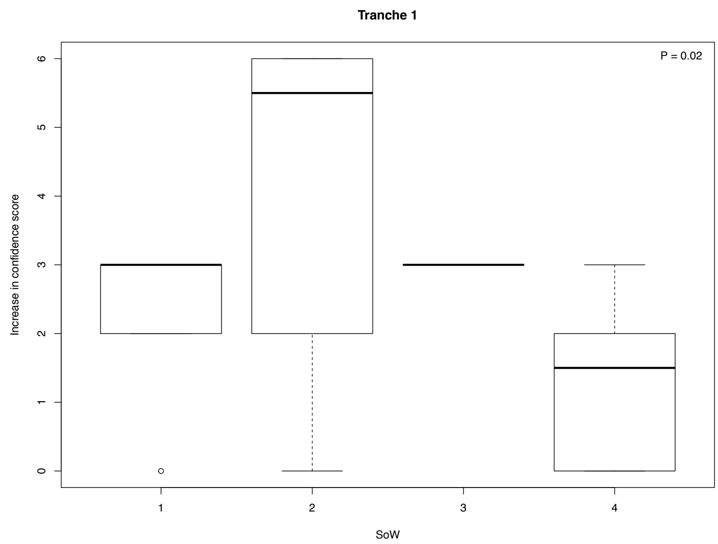
Note: a total of 11 factors were included in this multi-test correction. To be significant the adjusted P value ≤ 0.05

5.3.11 No evidence of biased distribution of teacher characteristics impacting on student level results.

At the student level repeatable evidence that all four Schemes of Work were all successful in improving student performance was found. However, two of the schemes were more effective (2 and 3) than the others, with Scheme of Work 1 being the least effective in both tranches. Could a biased distribution of teacher characteristics across the schemes account for these significant results? The data revealed no significant difference in the distribution of years of experience (χ^2 = 4.31, P = 0.23, Kruskal-Wallis rank sum test, tranche 1; χ^2 = 4.38, P = 0.22, Kruskal-Wallis rank sum test, tranche 2), completion bias (χ^2 = 7.03, P = 0.07, Kruskal-Wallis rank sum test, tranche 1; χ^2 = 4.84, P = 0.18, Kruskal-Wallis rank sum test, tranche 2), understanding (χ^2 = 1.94, P = 0.59, Kruskal-Wallis rank sum test, tranche 1; χ^2 = 2.07, P = 0.56, Kruskal-Wallis rank sum test, tranche 2) or acceptance (χ^2 = 2.27, P = 0.52, Kruskal-Wallis rank sum test, tranche 1; χ^2 = 5.45, P = 0.14, Kruskal-Wallis rank sum test, tranche 2) of evolution across the four schemes in either tranche. While the distribution of perceived increase in confidence level was not significantly different in tranche 2 (χ^2 = 2.70, P = 0.44, Kruskal-Wallis rank sum test), it was found to be significantly different in tranche 1 (χ^2 = 9.50, P = 0.02, Kruskal-Wallis rank sum test; Fig. 5.13) with a moderate effect size (ε^2 =

0.24). However, on closer inspection there was no significant difference between the distribution of confidence levels between Schemes of Work 2, 3 or 1, the only significant difference in confidence being between Schemes of Work 2 and 4 (P = 0.02, Dunn post-hoc with a Bonferroni correction), Scheme of Work 4 having the lowest distribution of teacher increase in confidence levels.

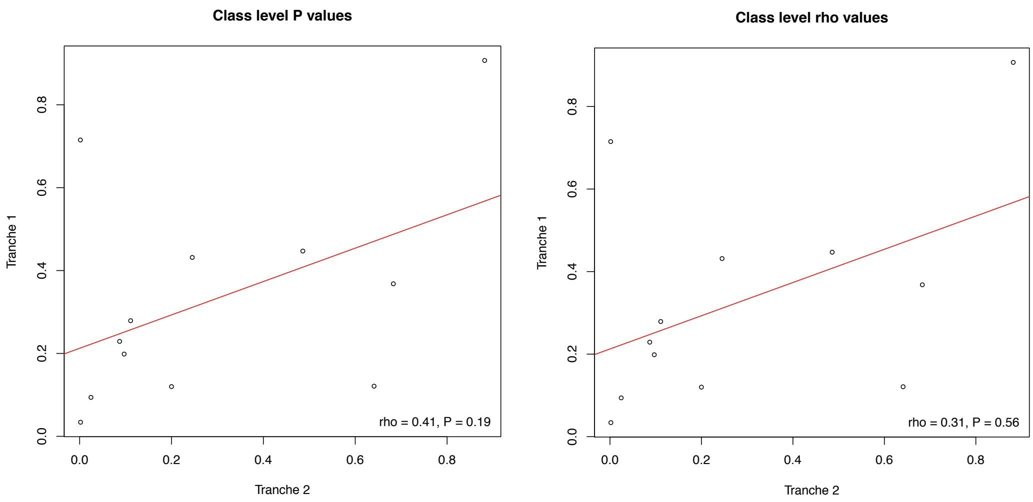
Figure 5.13 Boxplot showing relationship between the distribution of teacher increase in confidence scores stratified for Scheme of Work (Scheme of Work) in tranche 1.



5.3.12 Results of the class level analyses were inconsistent

If the class level factor results show reproducibility between data sets the values of P and rho taken from the individual equivalent class level statistical tests would be expected to show significant positive correlations. The data showed moderate positive correlations between P values (ρ = 0.41, P = 0.19, Spearman's rank correlation coefficient, n = 11) and rho values (ρ = 0.31, P = 0.56, Spearman's rank correlation coefficient, n = 6). See Fig. 5.14. However, neither correlation was significantly significant, probably due to small sample sizes.

Figure 5.14 Scatterplots with lines of regression showing correlation between P and rho values obtained from class level analyses in both tranches



5.4 School level analyses

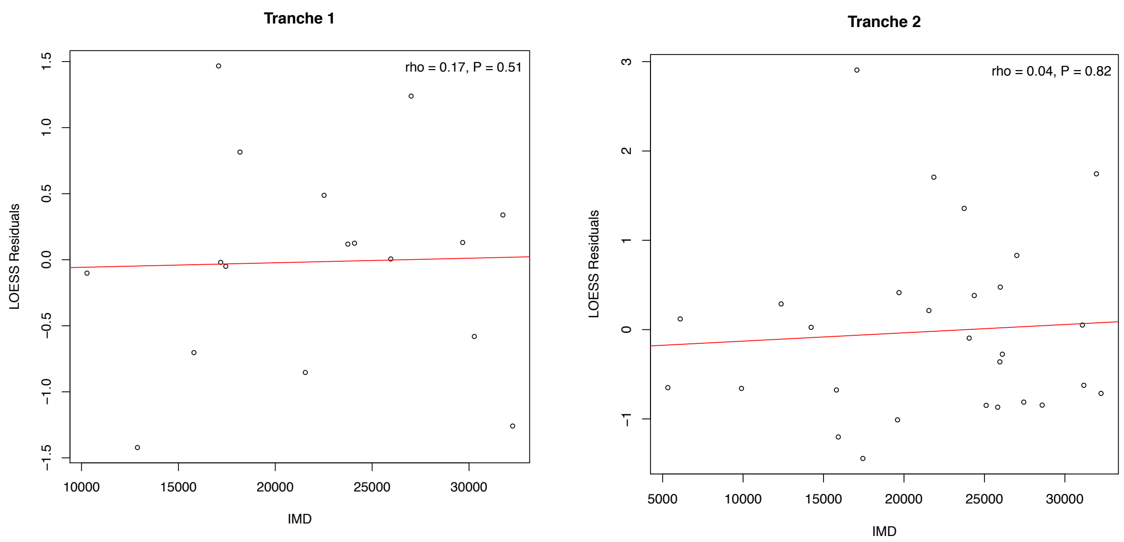
School level analyses were also carried out to compare performance between schools, using mean school LOESS residual scores (n = 17 in tranche 1, n = 28 in tranche 2) and data gathered from the school's most recent Ofsted publication / Ofsted website (https://reports.ofsted.gov.uk/) and data gathered from the English Indices of Deprivation 2015 (http://dclgapps.communities.gov.uk/imd/idmap.html). The English Indices of Deprivation 2015 are based on 37 separate indicators, organised across seven distinct domains of deprivation which are combined, using appropriate weights, to calculate the Index of Multiple Deprivation 2015 (IMD 2015). This is an overall measure of multiple deprivation experienced by people living in an area and is calculated for every Lower layer Super Output Area (LSOA), or neighbourhood, in England. Every such neighbourhood in England is ranked according to its level of deprivation relative to that of other areas. The Index of Multiple Deprivation ranks every small area in England from 1 (most deprived area) to 32,844 (least deprived area).

5.4.1 Relative measures of deprivation were not barriers to improvement in performance.

For the purpose of this study Index of Multiple Deprivation (IMD) together with 3 other domains of deprivation which could potentially affect the performance of participating children, IDACI (Income Deprivation Affecting Children Index), ESI (Education, Skills and Training Deprivation Index) and LEI (Living Environment Deprivation Index) were correlated with mean school LOESS residual scores. Values of the indices were obtained using the school's post code and therefore it was assumed that students attending a school lived within the catchment area of that school and within the same LSOA. Students within school populations from areas of higher relative deprivation would not be expected to perform as well as their less deprived counterparts.

In both tranches there was a slight positive correlation between increasing Index of Multiple Deprivation (IMD) scores (schools becoming less deprived) and mean school LOESS residual score in line with the expected trend. However, the correlations were not statistically significant with negligible effect sizes (ρ = 0.17, P = 0.51, Spearman's rank correlation, tranche 1; ρ = 0.04, P = 0.82, Spearman's rank correlation, tranche 2; Fig.5.15), with a combined P value of 0.78 (Fisher's test). These results suggesting Index of Multiple Deprivation (IMD) scores did not affect performance.

Figure 5.15 Scatterplots with lines of regression showing correlation between mean school LOESS residual scores and Index of Multiple Deprivation (IMD) scores in both tranches.



Similarly, positive correlations between increasing Living Environment Deprivation Index (LEI) and Income Deprivation Affecting Children Index (IDACI) scores and mean school LOESS residual score in both tranches were found which again were not statistically significant with small effect sizes. LEI (ρ = 0.39, P = 0.13, Spearman's rank correlation, tranche 1; ρ = 0.13, P = 0.51, Spearman's rank correlation, tranche 2; Fig. 5.16), with a combined P value of 0.24 (Fisher's test). IDACI (ρ = 0.14, P = 0.60, Spearman's rank correlation, tranche 1; ρ = 0.12, P = 0.55, Spearman's rank correlation, tranche 2; Fig. 5.17), with a combined P value of 0.70 (Fisher's test). The small positive correlations found with these three measures of deprivation (IMD, LEI and IDACI) followed the expected trend between improved performance with decreasing deprivation levels (i.e. scores got higher) however, none of the relationships were significant.

Figure 5.16 Scatterplots with lines of regression showing correlation between mean school LOESS residual scores and Living Environment Deprivation Index (LEI) scores in both tranches.

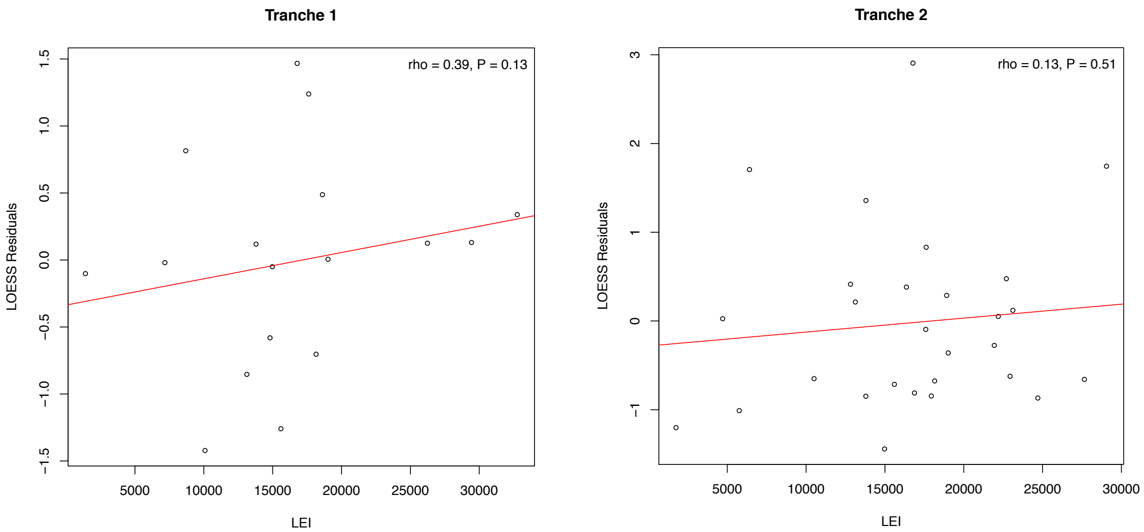
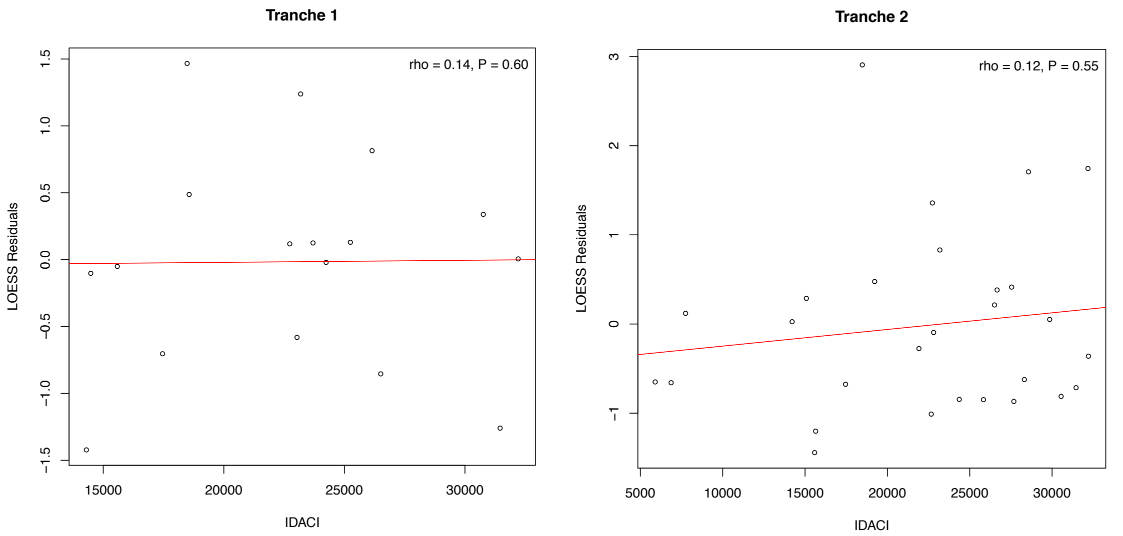
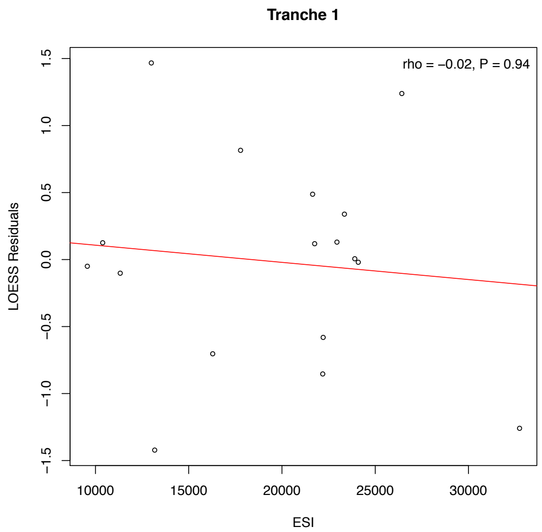


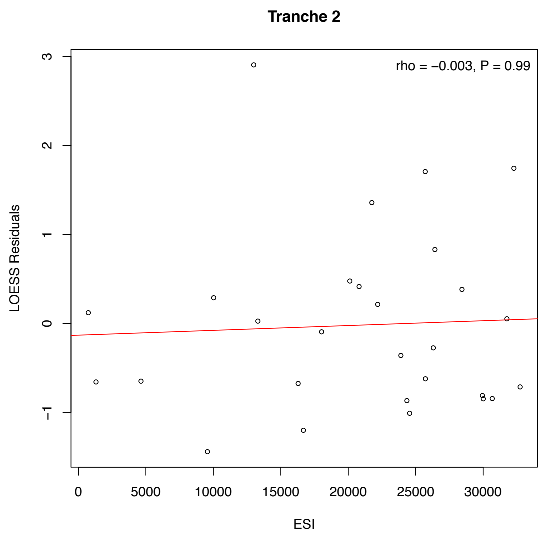
Figure 5.17 Scatterplots with lines of regression showing correlation between mean school LOESS residual scores and Income Deprivation Affecting Children Index (IDACI) scores in both tranches



Contrary to the trend expected there was a slight negative correlation between Education, Skills and Training Deprivation Index (ESI) scores and mean school LOESS residual score in both tranches suggesting that performance decreased as deprivation levels decreased. However, the relationship in both tranches reported small effect sizes which were not statistically significant (ρ = -0.10, P = 0.94, Spearman's rank correlation, tranche 1; ρ = -0.03, P = 0.99, Spearman's rank correlation, tranche 2; Fig. 5.18), with a combined P value of 0.99 (Fisher's test). This suggests that for the Education, Skills and Training Deprivation Index (ESI) deprivation domain performance decreased with decreasing deprivation but not significantly. Overall, these results suggest that the four measures of deprivation investigated pertaining to students within the school's population were not barriers to their performance.

Figure 5.18 Scatterplots with lines of regression showing correlation between mean school LOESS residual scores and Education, Skills and Training Deprivation Index (ESI) scores in both tranches

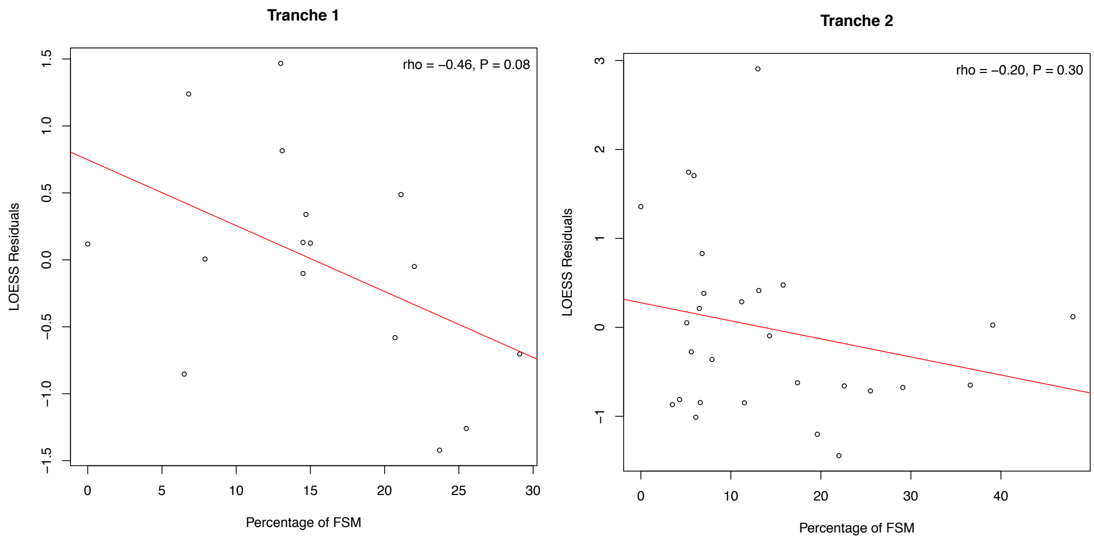




Note: In tranche 2 the line of regression is positive as it is produced by a linear model. When a Pearson's test was carried out on the data a positive correlation was obtained (r = 0.05, P = 0.80). The Spearman's test carried out on the non-parametric data is a non-linear test based on ranking.

The percentage of students qualifying for Free School Meals (FSM) was also considered as another independent measure of relative deprivation within the school population, as such we would expect a negative correlation between the percentage of Free School Meals and school performance. A correlation between Free School Meals and mean school LOESS residual score was carried out to verify the results of the indices of deprivation. In both tranches of data we found negative correlations between increasing percentage of Free School Meals and mean school LOESS residual scores were found which were not statistically significant (ρ = -0.46, P = 0.08, Spearman's correlation, tranche 1; ρ = -0.20, P = 0.30, Spearman's rank correlation, tranche 2; Fig. 5.19), with a combined P value of 0.11 (Fisher's test), and a medium effect size in tranche 1 and a small effect size in tranche 2. This result confirmed the expected trend of increasing percentage of Free School Meals with decreased school performance, however, this relationship was not significant and so suggests that this measure of deprivation was not a barrier to student understanding of this topic confirming the findings from the indices of deprivation.

Figure 5.19 Scatterplots with lines of regression showing correlation between mean school LOESS residual scores and percentage of Free school meals (FSM) in both tranches

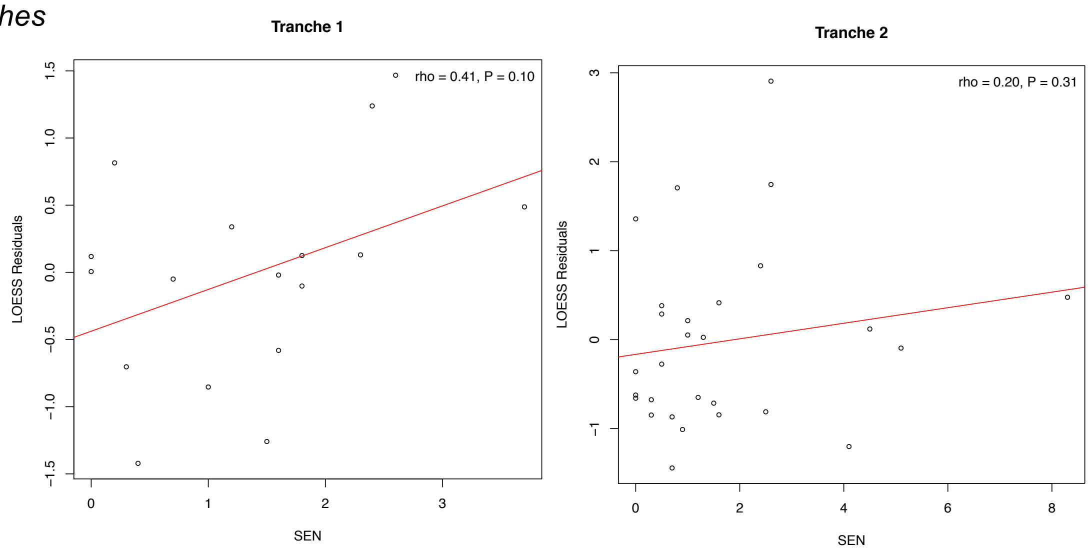


5.4.2 Factors relating to additional educational need were not barriers to understanding.

Two different measures of additional educational need were taken from Ofsted data, the percentage of students with Special Educational Needs (SEN) and those with English as an Additional Language (EAL). Both factors were expected to show a negative correlation with mean school LOESS residual scores as these cohorts of students should find it harder to understand the difficult concepts covered in the teaching intervention programmes.

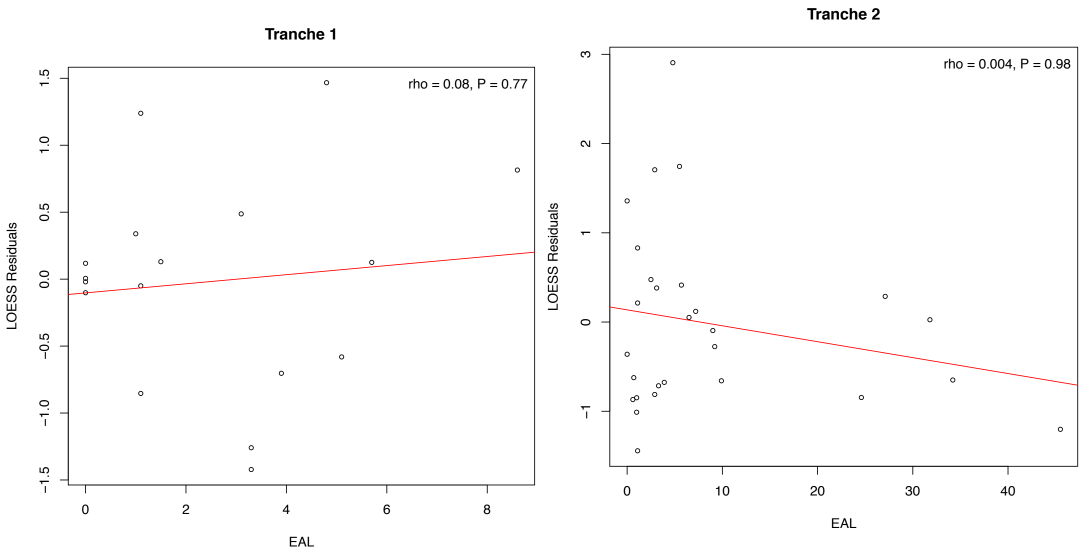
In both tranches positive correlations between increasing percentage of Special Educational Needs and mean school LOESS residual scores were found which were not statistically significant (ρ = 0.41, P = 0.10, tranche 1, Spearman's rank correlation; ρ = 0.20, P = 0.31, Spearman's rank correlation, tranche 2; Fig. 5.20), with a combined P value of 0.14 (Fisher's test), and a medium effect size in tranche 1 and a small effect size in tranche 2.

Figure 5.20 Scatterplots with lines of regression showing correlation between mean school LOESS residual scores and percentage of Special Educational needs (SEN) in both tranches



A similar but less pronounced trend was mirrored when the relationship between the percentage of English as an Additional Language (EAL) and mean school LOESS residual scores was correlated in both tranches. Slight positive, non-significant correlations with negligible effect sizes were obtained (ρ = 0.08, P = 0.77, Spearman's rank correlation, tranche 1; $\rho = 4.10 \times 10^{-3,}$ P = 0.98, Spearman's rank correlation, tranche 2; Fig. 5.21), with a combined P value of 0.97 (Fisher's test).

Figure 5.21 Scatterplots with lines of regression showing correlation between mean school LOESS residual scores and percentage of English as an Additional Language (EAL) in both tranches



Note: In tranche 2 the line of regression is negative as it is produced by a linear model. When a Pearson's test was carried out on the data a negative correlation was obtained (r = -

0.21, P = 0.28). The Spearman's test carried out on the non-parametric data is a non-linear test based on ranking.

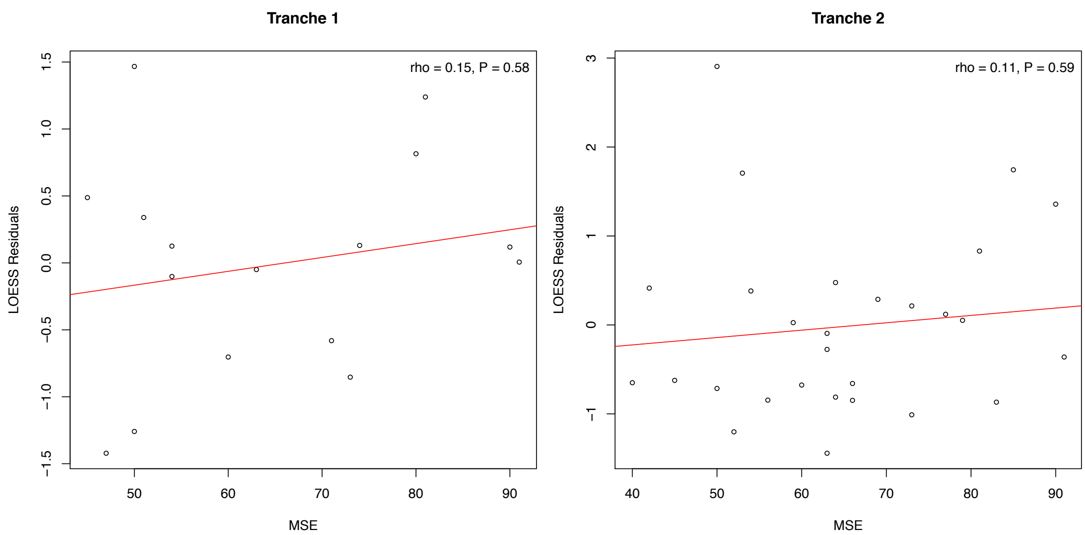
Together, these results suggest that students with Special Educational Needs or English as an Additional Language were not disadvantaged within this study, these factors failing to act as barriers to improved student understanding within the mainstream classes participating.

5.4.3 Factors relating to school performance were not predictors of improvement in understanding.

Two different measures of relative national school performance were taken from Ofsted data, the percentage of students meeting the standard level expected at the end of Key Stage 2 (MSE) and the school's most recent Ofsted rating grade. Both factors were expected to show a positive relationship with mean school LOESS residual scores with higher performing schools being more likely to employ high performing teachers who would be more effective in the classroom.

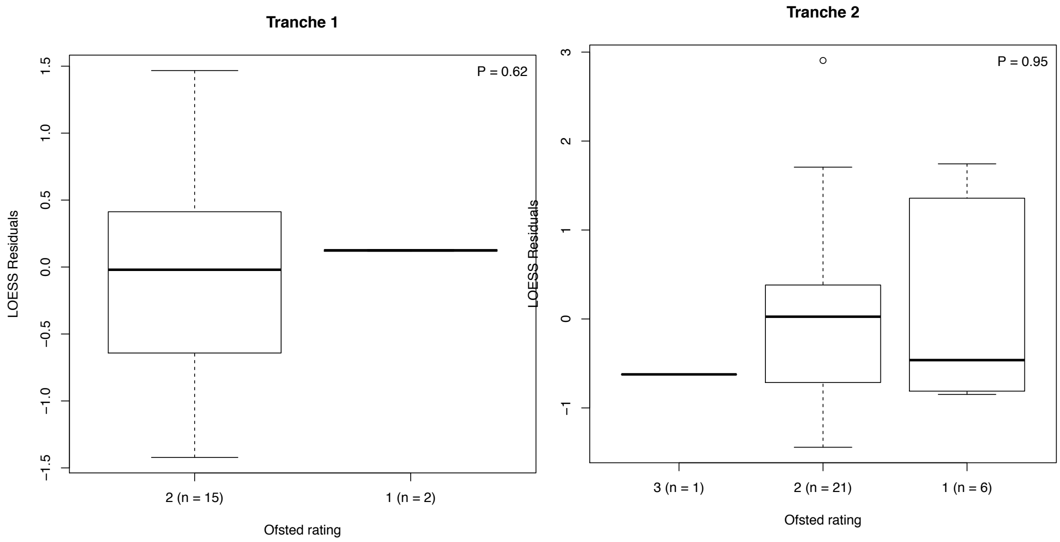
In both tranches there was a positive correlation between increasing MSE percentage and mean school LOESS residual scores which were not statistically significant ($\rho = 0.15$, $P = 0.58$, Spearman's rank correlation, tranche 1; $\rho = 0.11$, $P = 0.59$, Spearman's rank correlation, tranche 2; Fig. 5.22), with a combined P value of 0.71 (Fisher's test), and a medium effect size in tranche 1 and a small effect size in tranche 2.

Figure 5.22 Scatterplots with lines of regression showing correlation between mean school LOESS residual scores and the percentage of students meeting the standard level expected at the end of Key Stage 2 (MSE) in both tranches.



An analysis was also carried out to examine whether there was a relationship between the school's most recent Ofsted grade and student performance. Schools with higher Ofsted rating grades would be expected to promote a greater improvement in mean school LOESS residual scores. However, there was no significant difference between these two groups of teachers in either tranche (w = 11, P = 0.62, Wilcoxon rank sum test, tranche 1; χ^2 = 0.10, P = 0.95, Kruskal-Wallis rank sum tests, combined P = 0.90, Fisher's test; Fig.5.23), both with small effect sizes (Cliff's delta = -0.27, tranche 1; $\varepsilon^2 = 4 \times 10^{-3}$, tranche 2).

Figure 5.23 Boxplots showing relationship between mean school LOESS residual scores and most recent Ofsted rating grade in both tranches.



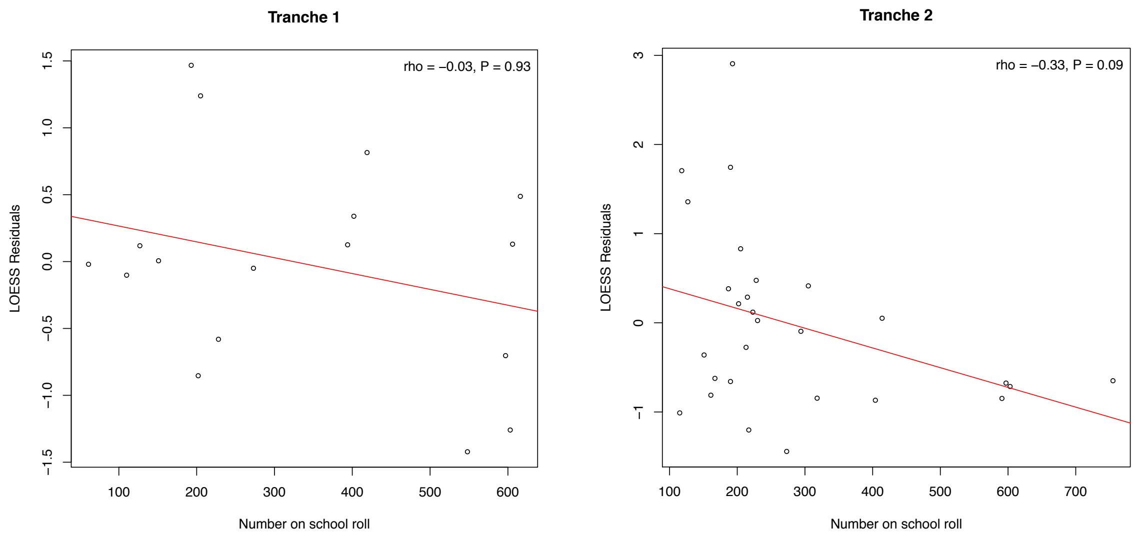
Note: Ofsted rating grades, 1 = outstanding, 2 = good, 3 = requires improvement.

5.4.4 Only the type of school attended was a predictor of improvement in understanding.

Five different school characteristics that could potentially impact on student learning were taken from Ofsted data to identify possible confounding factors. The type (primary or middle) and category (academy, managed, independent) of school, it's religious affiliation, size and teacher pupil ratio. The relationship between these factors and mean school LOESS residual scores were investigated.

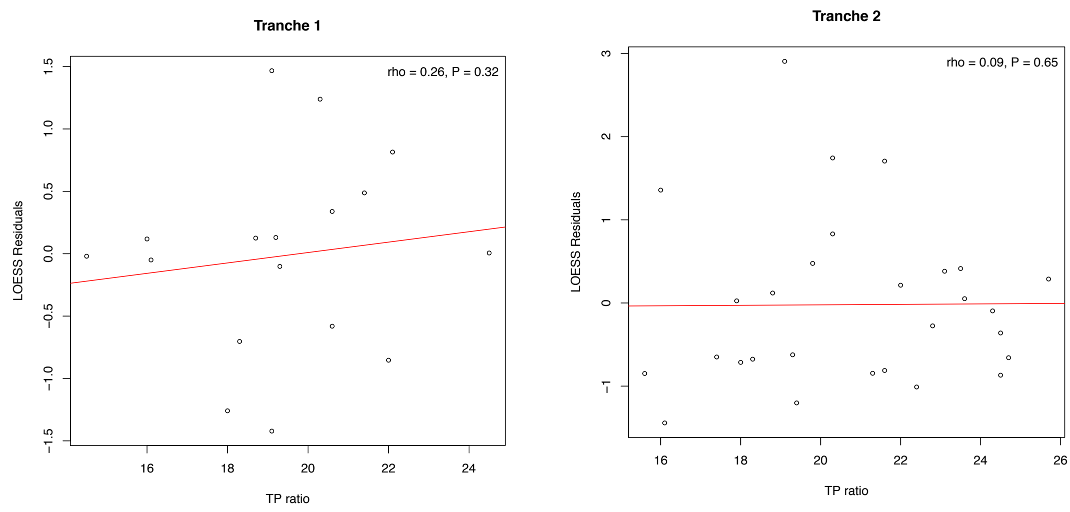
First, the size of the school was investigated to confirm whether it was an important factor. In both tranches there was a negative correlation between increasing number of students on roll (a measure of school size) and mean school LOESS residual scores which were not statistically significant (ρ = -0.03, P = 0.93, Spearman's rank correlation, tranche 1; ρ = -0.33, P = 0.09, Spearman's rank correlation, tranche 2; Fig. 5.24), with a combined P value of 0.98 (Fisher's test), a negligible effect size in tranche 1 and a small effect size in tranche 2. These results suggest that the size of the school is not a predictor of student performance.

Figure 5.24 Scatterplots with lines of regression showing correlation between mean school LOESS residual scores and number of students on school roll in both tranches.



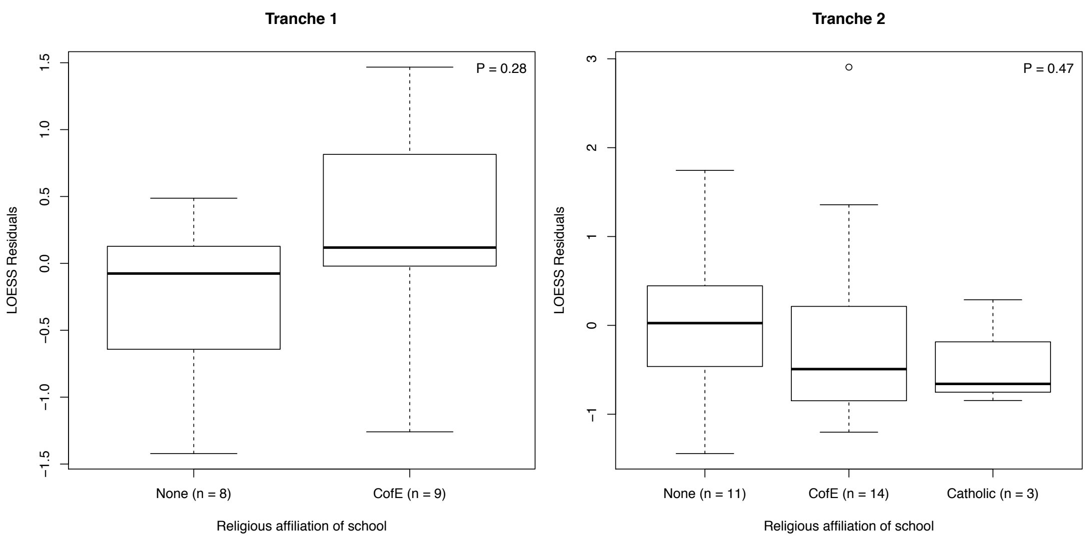
Schools with lower teacher pupil ratios might be expected to be more effective as a greater proportion of teaching time could be devoted to each individual within the class. However, the results show an unexpected positive correlation between increasing teacher pupil ratio and mean school LOESS residual scores. These correlations were not statistically significant (ρ = 0.26, P = 0.32, Spearman's rank, tranche 1; ρ = 0.09, P = 0.65, Spearman's rank correlation, tranche 2; Fig. 5.25), with a combined P value of 0.54 (Fisher's test), with small effect sizes in both tranches, suggesting that teacher pupil ratio is not a barrier to student performance.

Figure 5.25 Scatterplots with lines of regression showing correlation between mean school LOESS residual scores and teacher pupil ratio in both tranches.



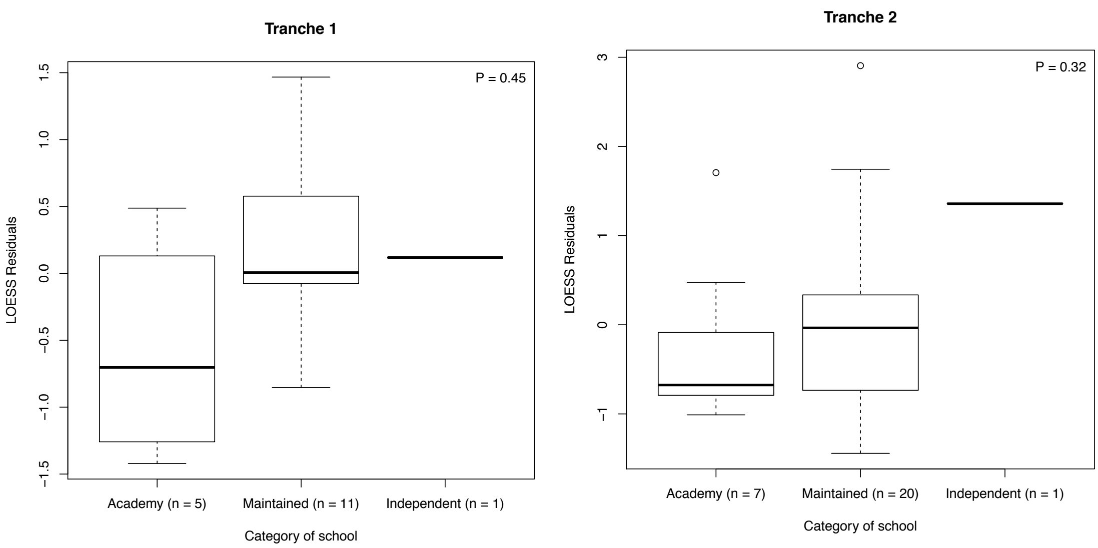
While, church schools have been reported to be more effective than more secular schools they might also experience more difficulty in teaching evolution; does a school's religious affiliation impact on student performance? The results showed no significant difference between mean school LOESS residual scores and their religious affiliation in either of the tranches (w = 24, P = 0.28, Wilcoxon rank sum test, tranche 1; χ^2 = 1.49, P = 0.47, Kruskal-Wallis rank sum test, combined P = 0.40, Fisher's test; Fig.5.26), with a medium effect size in tranche 1 (Cliff's delta = -0.33) and a small effect size in tranche 2 (ε^2 = 0.06), suggesting that a school's religious affiliation was not a predictor of student performance.

Figure 5.26 Boxplots showing relationship between mean school LOESS residual scores and religious affiliation of school in both tranches.



The category of a school could also affect performance due to the effect of differential funding. However, there was also no significant difference between the different categories of school and their performance when analysed against mean school LOESS residual scores in either tranche (χ^2 = 1.6, P = 0.45, Kruskal-Wallis rank sum test, tranche 1; χ^2 = 2.3, P = 0.32, Kruskal-Wallis rank sum test, combined P = 0.42, Fisher's test; Fig.5.27), both with medium effect sizes (ε^2 = 0.1, tranche 1; ε^2 = 0.09, tranche 2). This suggests that a school's category was not a predictor of student performance.

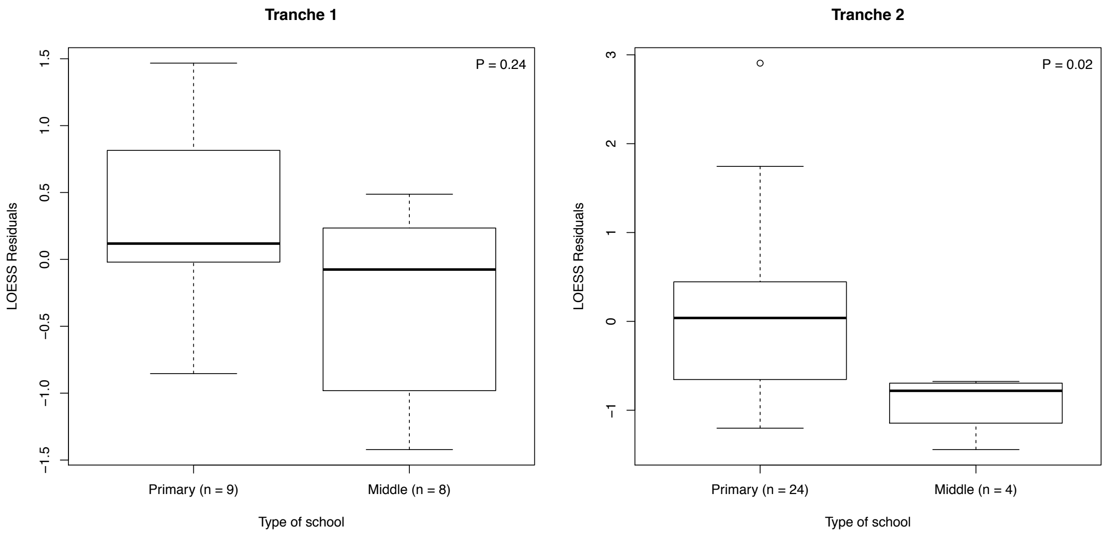
Figure 5.27 Boxplots showing relationship between mean school LOESS residual scores and the category of school in both tranches.



Finally, type of school attended (either primary or middle) was investigated to determine whether this could affect student performance. Teaching of the topic in middle schools might be expected to be better as they have more science specialist teachers or conversely better in primary schools, where there is greater lesson time flexibility and the teachers know their students more intimately. When a correlation between the type of school and their mean school LOESS residual scores were compared no significant difference between them in tranche 1 (w = 49, P = 0.24, Wilcoxon rank sum test) was found, with a medium effect size (Cliff's delta = 0.36). However, in tranche 2 primary schools performed significantly better than middle schools (w = 83, P = 0.02, Wilcoxon rank sum test) with a large effect size (Cliff's delta = 0.73). See Fig. 5.28. These results giving a significant combined P value (P = 0.03, Fisher's test) suggesting higher student performance in primary school settings compared to middle schools. In light of this result and to check the validity of our findings at the student level, an additional set of student level analyses were carried out excluding all data obtained in middle schools. As this significantly reduced the sample size the results of both tranches were combined before analysis (n = 1249, n = 313, tranche 1; n =

936, tranche 2). Interestingly, the results of the new analysis excluding middle school data were very similar to those of the whole tranches; confirming a negligible non-significant correlation with age (ρ = 1.60x10^{-2}, P = 0.57, Spearman's rank correlation), no significant difference with gender (w = 201120, P = 0.34, Wilcoxon rank sum test) with a negligible effect size (Cliff's d = 0.03), a small significant positive correlation with ability (ρ = 0.20, P = 1.15x10^{-12}, Spearman's rank correlation) and a significant difference between the different Scheme of Work (Scheme of Work) (χ2 = 26.93, P = 6.088x10^{-6}, Kruskal Wallis test) in which Scheme of Work 2 and Scheme of Work 3 were the most effective.

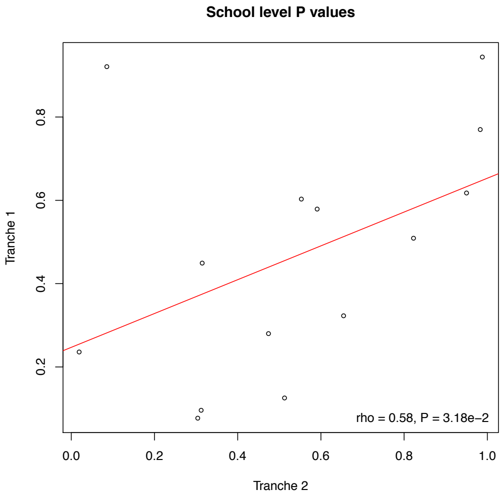
Figure 5.28 Boxplots showing relationship between mean school LOESS residual scores and type of school in both tranches.

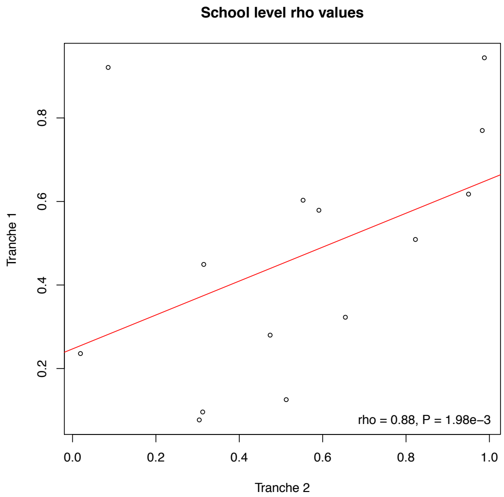


5.4.5. Evidence of reproducibility of results in school level analyses

If the school level factor results show reproducibility between data sets it would be expected that the values of P and rho taken from the individual equivalent statistical tests would show significant positive correlations. The data showed significant strong positive correlations between P values (ρ = 0.58, P = 3.18x10^{-2}, Spearman's rank correlation coefficient, n = 14) and rho values (ρ = 0.89, P = 1.98x10^{-3}, Spearman's rank correlation coefficient, n = 10) obtained in the individual school level tests carried out to identify possible predictors of performance, the consistency of these results showing evidence of repeatability at this level of analysis. See Fig. 5.29.

Figure 5.29 Scatterplots with lines of regression showing correlation between P and rho values from individual statistical tests in both tranches





5.4.6. Summary of school level analyses

Although one significant result from the analysis of school levels factors was obtained; primary schools out performing middle schools (Combined P = 0.03, Fisher's test) with a large effect size in tranche 2 (Cliff's delta = 0.73), it should be noted that this one significant result failed to meet the correction for multiple tests with an adjusted P value of 0.40 (Bonferroni or Holm) for the 14 factors tested. Overall, the data revealed repeatable evidence that none of the factors analysed were significant predictors of performance at this level of analysis, suggesting the school-based factors investigated did not hinder the teaching of this topic; all schools increasing the understanding of their students.

5.5 Chapter conclusion

The results suggest that whilst the teachers in this study showed a high acceptance of evolution they did not have an adequate understanding of the process of natural selection, their alternative conceptions mirroring those reported in their students post instruction. However, their effectiveness was not affected by either of these two factors. In the teacher level analyses only three significant factors relating to teacher confidence were identified; years of experience, perceived improvement in confidence and completion of the pre-teaching questionnaire, whilst at school level, only the type of school attended (either primary or middle) was a significant predictor of performance with evidence of repeatability of results. Overall, the findings from this chapter suggest that only these four factors significantly influenced teacher effectiveness from the wide range examined, although all failed multi-test correction tests. In the next chapter the methodology of qualitative data collection is discussed and a representative summary of the participating teacher feedback is shown.

Chapter 6 Qualitative teacher feedback

6.1 Chapter overview

This chapter outlines the methodology employed to gather and code the qualitative feedback obtained from the teacher interviews and fieldnotes taken during the post teaching interviews conducted in all schools. Representative teacher comments from the emergent themes are grouped together to add richness and help to explain some of the results obtained in the quantitative analyses.

6.2 Methodology employed to collect qualitative teacher data

Qualitative data were collected using semi structured interviews (see Appendix J3) in all participating schools post instruction conducted in a pre-arranged mutually convenient time slot of at least an hour. Over 60 hours of data was collected from audio taped interviews with the class teacher or small focus groups (where there was more than one teacher) together with fieldnotes taken at the time. The audio recordings were transcribed verbatim professionally focusing on the informational content of the discussion or ‘denaturalized transcription’.

All transcriptions and other textual material were then read and re-read by the principle researcher to ensure familiarization with their content. Systematic, thematic emergent coding was then carried out. This process involved coding the data by identifying important themes and phrases, then isolating emergent patterns by looking for commonalities and differences within each theme and then relating any consistencies to the quantitative data.

Data was hand coded by reviewing all data line-by-line, identifying key issues or themes and then highlighting segments of text (either original text or summarized notes) to those themes in different colours. New themes were added as additional themes or issues emerged in the data. The data was coded comprehensively (i.e. no original data remain un-coded), coding segments of text to multiple themes if a single statement contained information relevant to more than one theme. Data within each identified theme were then isolated form the main body of the text and collated.

Data from each isolated theme was then read and re-read in order to create a hierarchical ‘tree’ broader descriptive themes were broken down into smaller coding units or sub-themes. Each sub-theme was then read and reread in order to make comparisons between this smaller set of isolated text passages. Passages within sub-themes containing similar points were nested together and contrasting points were also identified. Due to the volume of data generated by the interview process representative points were identified within each sub-theme to be included in the thesis to illustrate the diverse opinions of the participating teachers. (Please note that comments pertaining to the relative effectiveness of the different main activities in lesson 2; hunting moths or a power point with scaffolded task and lesson 4; identification of homologous structures and common ancestry in trilobites or extant mammals, have already been explored in chapter 3).

Rather than identifying statistical differences within the qualitative data, the aim of this process was to look for clear patterns in the data which could be used in a complementary way to enhance and clarify the results of the quantitative analyses (Greene et al., 1989) regarding the utility of the teaching resources and testing regime developed by the study. These qualitative data were collected to add richness (Farber, 2006) and to strengthen the research findings by trying to "capture the complex, multifaceted aspects of teaching and learning" (Kagan, 1990, p. 459). The comments that follow are grouped into the themes and sub-themes obtained from the coding process. The comments included in this thesis show the range of similarities and differences in teacher opinion and were carefully chosen to be representative of the whole larger data set.

6.3 Teachers gave feedback on the utility of the student assessment instrument

The teachers gave their opinions on the suitability and utility of the student assessment instrument employed by this study.

6.3.1 The student questionnaire was perceived to be too hard

There was unanimous agreement between the teachers in the post teaching interview that the student questionnaires were too hard with several different contributing factors identified:

1. The students were unfamiliar with the style of questionnaire assessment items and their method of delivery

"It's not something that they're used to, they're not used to that style, although they do have multiple questions on some of their SAT papers, but of course they haven't had SATS yet,

and neither have we started doing practice papers per se, so they rarely come across that format of questioning".

"The questionnaires took quite a lot of time, because that's not really how teach, standing at the front saying here's a question, write it down, the answer, because like you say, they couldn't remember the start of the question, some of them went over more than one PowerPoint slide and some of them were quite hard to understand".

2. The assessment instrument language level was felt to be too high and too verbose which could have demotivated the students leading to lower levels of performance.

"I mean that was sort of the low point in that also, you know, if you look at some of, I mean I know that you are obviously testing what they know and their vocabulary, but some of it was really, really wordy and even as adults when we were sort of looking through it, we were sort of thinking, 'What is that?' It was quite difficult trying to unpick, so my sort of concern with if you actually manage to get any, you know, quantitative data from it because, you know, I was saying to them, you know, "If you really don't know, you're just going to have to have a guess".

"*There was a lot of information for them to retain as they were, as you were reading it through and for some of them, you know, more than ten words and they're actually forgotten what the question was".*

"The language for the questions were very, very difficult for them, very long and that was another problem that we didn't almost have the time to do these. The first time it took an hour, almost an hour, and they lost the will to live".

"Yeah, the questionnaires were too hard, because they were doing them, they kind of looked, on almost every question, they looked at us blankly, and by the time we'd got to the end of the question, they'd forgotten the beginning of the question. I mean we're dealing with pupils, some of who have got reading ages of a six-year-old and some have got 15. Some of the language was really quite hard for them to understand with like the common ancestors and just all that, because it was kind of just going straight into it. So, overall, I think it was pitched a bit too high, it was too hard".

"I think the opening one [questionnaire] almost put them off the survey, they're looking even the sort of brighter, um, um, um, what are you asking me to write here?".

"…and it was quite wordy, that took quite a long time to get through all the information so I think those with a shorter attention span may have just…turned off, turned off, really".

3. The alternative response options within each assessment item were felt to be too similar and confusing.

"We found the questionnaires too hard, it took us an hour to go through the questionnaire and we found that children that knew the answer, once they'd read the four, they didn't know the answer, they didn't know which one to tick because it was quite wordy and so in the end it was tick one because I think it will be a 25% right answer, I think it will be the one you thought because they couldn't access that level. For both classes and they lost the desire to participate after about question six".

"Yeah, my groups did struggle with the questionnaire because it was quite a lot of reading involved, and listening as well. They found them (assessment item options) very similar".

"Yes, you have to read them at least twice to kind of get your head around, what is this asking? Because the options A, B, C, D, they're quite lengthy and quite similar, you have to read those a couple of times".

4. Some teachers worried that the assessments wouldn't reflect the actual progress made by their students based on their observations during teaching the topic.

"I don't know whether the student questionnaires then show what they have really learnt because they struggled a lot with the amount of reading. I mean, I read it to them. We had it on the whiteboard and I was reading it all to them, but just taking in all the information for each one. I think through the lessons I think they've learnt an awful lot, but I don't know whether it's going to show hugely in the questionnaires".

"I don't know what the questionnaires will show, but my teacher observation would be that there was a huge progress that was made and you can see from lesson to lesson that they were starting to apply stuff from the lesson before the birds and the beaks and they were starting to kind of bring that in".

“They found the multiple-choice questions hard, they found actually holding onto the question very, very difficult. I just didn’t want you to feel that it hadn’t been successful because I felt they had made so much progress”.

5. The pitch of the lessons and the assessment was mismatched

“The content was at the right level but the way that the questions were asked didn’t match the content, they were much too hard”

6. One teacher even reported that their students felt stressed when completing the first questionnaire.

“I found that as I was reading it, I was having to go so that one’s saying that and that’s saying that, to try and pick out which one was which, or saying it’s this or it’s that. That was the worst thing that we had to do, they really didn’t like it. Especially the first time, they were dreadful, they hadn’t got a clue and they felt quite stressed.”

However, the need for a discriminating assessment instrument was appreciated by the teachers, with all teachers recognising that the second and subsequent assessments were quicker and easier with evidence of greater understanding from observations made over the course of later assessments.

“I know that the whole point was to show progress between what they understood at the start and what they understood at the end, but that first questionnaire was really difficult for them”.

“They found the first part of the questionnaire very difficult. When we did the original one they had no idea for this bit really at all. They weren’t too bad with the second one which only took about 20 minutes. Actually, the second time round they were absolutely fine with doing that and they definitely found those questions easier”.

“The first one took a long time, and a lot of step by step to read out each part, but when we did them again a lot of them were like, "Oh yes, I know now." That sort of comment, "I now know what the answer is." Rather than before they were just guessing something, this time a lot more of them were saying, "I know this, that makes more sense. Last time I put this," and kind of laughing at their last answer they thought they'd put”.

"I would say that they found, when they did the questionnaire at the beginning, they found it really, really difficult, and I didn't know the answers to some of them either. They found it really challenging, so lots of them were just completely guessing, but then when we did the questionnaire after the unit of work, they all did loads better, and they actually could access the questions. So even if they didn't know the answer, they understood what the question was asking. So that was good, it shows that they've learnt something".

"The questionnaires were really hard the first time around. But it was quite interesting when they did the one the second time, some of them just went oh yeah, and that was it. Yeah, and it was sort of, oh right, you know, and it wasn't a big deal. Whereas the first time around it was quite arduous".

"The questionnaires were tricky, a bit tricky. But, by the time we did it the third time, they were so used to it, it was easy and much quicker to do".

6.4 Teachers gave feedback on the utility of the resources and learning within the classroom

The teachers gave feedback on the teaching resources and different activities contained within the Schemes of Work (SoW).

6.4.1 Studying variation was engaging with good links to maths skills

Teachers in the study felt that the main activity from lesson 1: looking at variation within the class, was a worthwhile and appropriate activity. They particularly like the practical element of the lesson which did not need heavy resourcing.

"Oh yeah, comparing their handspans, the practical was engaging, it wasn't just look at this graph and interpret this data and tell me, they really liked the actual measuring and doing stuff. That was really nice and easy, low resource. You have a hand, please stretch it out and measure it".

They also appreciated the links with numeracy and the way in which the activity could be used in a cross curricular way in order to 'shoehorn' some extra science into the timetable.

"The hand span exercise was good; especially the measuring and recording and its links with maths. And that was good where we had some maths within the science lesson".

“I think the children have a better understanding this year than they did last year without your project. Because for me, trying to fit in with the curriculum requirements in primary anyway, you always find that it's [science] been bumped out. Actually, science is really important, it's important they get their statutory allotted amount of time, and you have to link it in through other subjects, but often that loses that practical element. Because you end up doing it through literacy and science through maths, so you're doing bar charts, but you don't often do the practical side of things at the same time. Often science has to be heavily resourced, and we don't always have those resources available, or if they are, there’s not enough to use in small groups”.

“Yes, we linked it to the maths and we've been working out averages, so we looked at those sorts of things”.

“It's just trying to think about how we could actually make those links between maths and science. We've got, the way maths is taught now, going unit by unit, so thinking right let's do all the measuring and then the calculation in science. Actually, this activity lends itself really well to thinking, okay let's just do all science where we can look at our data and those sorts of things so that was nice. They really liked looking at all sorts of variation, cheek dimples and freckles”.

They liked the lay out of the activity sheet that provided scaffolding for the students enabling access by all abilities and its ability to identify the students needing extra help plotting the bar charts.

“I liked this simply because it did really help because quite often just drawing a table for some of them, especially my able groups, who, yes okay they might be great at Maths, but actually in terms of writing, it can be very slow sometimes. So, it was really great just to be able to get on and they could just fill in the tables”.

“The sheets that you gave us were really good in the fact that it showed them how you might lay it out. There were some children, that just stuck that in, because that was pitched right for them in terms of, they would have then struggled to draw more from that and to explain. So, for those children, they kept that and stuck it in their books, others I said, "Right, that's your plan, write that up now, what conclusions can you draw from that, is there anything else you can link in?" So that they can start pushing those other ones, that might have finished sooner. Yes, it acted as a scaffold, but then some others might be able to do more and take it further”.

"I think it highlighted children who are developing anyway in the section found that easy, but it highlighted that there were some skills we needed to reiterate in numeracy and that sometimes, some children needed a bit more emphasis with the modelling with how to do a bar graph. It showed who had mastery and who didn't because they could just go off and get on with it and the others you needed to give a little bit more help and scaffolding to."

Finally, the activity drew attention to the occurrence of intra specific variation allowing progression on to the next lesson on natural selection by differential survival and reproduction.

"The whole idea of variation just within one species was funny, because it's one of those things that is blindingly obvious and then you start talking about it and it's like, "That's why it's blindingly obvious." It was a nice way to start, because it's quite non-threatening. They all went home and said, "Can you do that?"

6.4.2 The 'Blob' game was enjoyable and helped to introduce the concept of natural selection

The 'Blob' game (part of the starter activity in lesson 2) proved to be a very popular activity to introduce the concept of natural selection acting on random variation within a population. Not only did the students enjoy it and want to play it again but they were able to understand and explain the underlying concepts:

"I'd say that they all enjoyed the activities, especially the active stuff, they really liked the Blob game that we played, they really enjoyed that."

"They really enjoyed that [the Blob game]. That one they wanted to do again, and again, and again, and again. I think we did we did a couple of lessons later, they wanted to do again."

"That was the starter bit, they enjoyed that. It was like "Oh, I've got to sit down now." They were desperate to play it again. They were like "Let's play this game. I want to see if I can survive." When they played it again they kept saying, "Can I have that one this time?" No, you can't have that one. It's got to be random. I think they definitely got that it was random."

"They did get what was going on. They got... you know, if they went out, they could tell you why they'd gone out, sort of thing – 'I drowned because I didn't have wings,' or whatever it

was; I can’t remember now. I know you... ’cause you can fly...’cause there wasn’t any land to land on...

“It was a lot of fun, that one.”

6.4.3 Time line activity illustrated the vastness of geological time

The schools reported that the time line activity (main activity in lesson 3) to introduce geological time was very successful in engaging student interest and particularly liked the visual practical nature of the lesson.

“It [time line] was the best thing out of the lot, the kids really loved it”

“Absolutely, it was really good, we went outside on to the playground to do the toilet roll timeline activity, which the children really enjoyed. So, I made sure I set really clear expectations in the classroom before I went outside. It really helped them to understand it better to visualise it rather than just reading words on a page.”

“The time line was fantastic. I loved that. It was really good fun visual activity.”

The teachers also reported that the activity was a really good introduction to the huge scale of geological time, a concept that the majority of their students had little understanding of before instruction.

“They enjoyed that. That was very good. We went out on to the playground with our loo roll and put the bits along. It really opened their eyes to how many years, how many millions and billions of years we weren’t on the planet. My class really loved that, really loved that.”

“The toilet roll, that was really good. They had no conception, none at all. They were completely shocked, well most of them, about 90% I’d say.”

“When they [the cards] were placed on that line and we tried to get a panoramic photo with them all in, we had to have it in a horseshoe because it wouldn't fit straight in the hall. We all struggled to grasp how long a period of history that is.”

“They really liked the toilet roll time line that we did. That, I think, was quite a powerful, visual, physical representation and they found that quite helpful.”

"They liked that [the toilet roll activity]. It blew their minds. It absolutely blew their minds because when they had a go first didn't they didn't have a clue with the non-dated cards. It was as if they'd just thrown them on the floor".

They reported that the toilet roll time line served as a memorable activity which their students were able to refer back to later in the topic.

"*Giving them a different and memorable task to do was very clever, when else are you going to be standing next to a toilet roll timeline? Then they can link the facts that they got from it, by remembering what they did."*

"Yes, they're all doing something, and I think also their understanding of geological time and the events as well was really good, it was so vastly different from anything else they've done."

"The timeline was fantastic, so I've got pictures of all us out and that, in fact the staff that came out with me were saying, "I've learnt so much, didn't have a clue." Just the visual aspect of it saying this is the beginning of the world, actually everything they've previously known as a timeline was all in that bit, I think that was really, really effective."

The teachers also appreciated the way in which the time line portrayed the uneven distribution of important events over the huge expanse of geological time as well as its links to numeracy and collaborative learning.

"I think they could see how long it took, because obviously there were huge gaps. They got the idea. It was such a long time ago."

"The toilet roll timeline was amazing, because so much happened in the last 5 sheets, nothing happened for the other 150, and then you get to the last 5 sheets. When they first put them down before it had the numbers on, they were so desperate to evenly space them."

"They really enjoyed it and I think what it is very useful for with that is getting a real sense of the fact that it's not just a few years ago, it's millions of years ago. I think for getting a scale it's fantastic, and I think possibly if I were to do it again in the future I probably would even do a big timeline across the classroom, so they could see the pictures and how it all fell into place."

“I think for some of them, yes, it gave them an idea of the vast distances between anything happening really, and how a lot of things happened, although with millions of years in between them, happened relatively quickly.”

“They were fine with that [ordering events] we've got some really strong mathematicians in the group, so that didn't present any issues. Then again it was the whole group, if we'd done it on a smaller scale then I think you would have seen the children struggle with it.“

“*It's really interesting, because obviously some of them have a much better conception of the numbers than the others. So, when I was putting them in groups to put the cards down I made sure there was somebody who had a conception of, it's before year zero, it's going back that way and the numbers are going to get bigger as we go that way. When I first put them down before it had the numbers on, they were so desperate to evenly space them.”*

“Again, it pings back to maths, because they've got to go up to 10 million or whatever it is, ordering and rounding. It is good to try and tie those in to those big numbers.”

Some teachers went further to endorse using the toilet roll time line activity to depict different events in history, suggesting that it was easier for their students to understand and more visual than using a clock face model.

“*The toilet-roll of time was a big hit. But I actually think it’s a really nice... You could adapt that to other things; other timelines in history. I think, with the toilet-roll, you see it, It’s more visual. With the clock...I find the lower-ability children struggle to see it as easily with the clock-face. And they can’t tell the time anyway, can they? So, it’s really hard. So, it’s much, much better with the toilet-roll.”*

However, a few teachers reported some logistical problems with the toilet roll timeline activity ranging from behavioural issues, the weather and the preparation time needed to set it up:

“*I didn’t manage that [time line] very well with my difficult group, their behaviour in the hall stopped me doing it. They were in corners and they were getting distracted. I didn’t have any support in that group as well so it was quite difficult but with my other higher ability group they l loved it and it was a really lovely activity. Because I could be in more places at once. I didn’t have to constantly look for trouble.”*

"It was quite difficult to manage that activity with the whole group, keeping everybody engaged on what we're doing."

"I found this one the trickiest out of all four, just because there wasn't an activity for them to work on independently. It was very much having to keep their attention and keep them focused in a group, which with this lot is quite tricky."

"The toilet roll timeline, which had been really successful last time, sadly it was a rubbish day and it was windy, and we came outside, we couldn't get in the hall, so that ended up not being as successful as it potentially could have been."

"I found the timeline, the toilet roll activity tricky because in terms of setting up time and only having me in the class at that point. I just found that really tricky to manage in terms of preparation time and keeping them on task."

To try to resolve some of these problems an alternative method of carrying out the time line activity was provided and used by some teachers who pre-emptively identified potential behavioural issues within their classes. This activity involved students working independently in small groups ordering and placing the same key events along smaller (4.60 m) multiple washing lines of time in the classroom.

"*The washing lines worked well. They enjoyed that and that was much easier to do than the toilet roll like we used last year. It was being able to work in smaller group and I had an LSA with the one group, helping them along."*

6.4.4 The teaching interventions were successful in introducing and reinforcing the concept of common ancestry

It was evident form the feedback that the teaching intervention packages were successful in introducing common ancestry and reinforcing this cognitively difficult concept throughout the sequence of lessons with both the teachers and the students. It should be noted that most of the comments about common ancestry were given during discussion of the time line activity in lesson 3 or studying the pentadactyl limb in lesson 4, very few comments were obtained from the feedback on Trilobite homology. A possible explanation for this was that the teachers observed the students making the link between homologous bones in their own arms and the other mammals on the sheet in a more obvious and concrete way than was possible in the Trilobite activity.

“Definitely, I’ve learnt loads. I’ve never taught common ancestry before. They [the students] were fascinated actually, fascinated by the fact that it [all life] all comes from that one organism.”

“It's the common ancestor, and it's funny, because from never saying it, that's probably the words we've said more since January than anything else. Like the tree and the lizard question I would say they all found it tricky the first time but at least half of them knew straight away the second time that of course they're related, it's the common ancestor. They didn't even have to think about it.”

“It was very successful, it was very exciting at times, and hearing how they were speaking and what they were coming out with and the conclusions they were drawing, like, oh he might be my distant cousin then because we’re all coming from the same thing, we’re related.”

“They quite liked the timeline one. They thought that was quite cool. They were very surprised. I don't think they realised we all came from one common ancestor. They took it for granted that we came from apes and that's they heard. I don't think they ever paused to think where the apes came from and how that all worked its way through. That was quite interesting and we had some interesting discussions”.

“And then it was the toilet-roll, which went really well. And I just feel, in terms of their retention of knowledge as well from that, I feel that was probably one of the strongest things that we did. They really did understand it and they referred to it themselves. When they independently bring things up, you know it’s really sunk in. Yeah, and it’s interesting you can keep coming back to it, because I do remember when we went on to common ancestors, that I could keep almost pointing at the point in the room. I said, ‘Do you remember that single cell that was there?”

6.4.5 The resources had good links to literacy

The resources were planned to teach evolution in a cross curricular way with intentional links to numeracy and literacy as key components in Key Stage 2 teaching and assessment. Optional exercises were developed to introduce three historical figures (Charles Darwin, Mary Anning and Jean Baptise Lamarck). The delivery of these comprehension activities was flexible allowing them to be used in a variety of ways. Some schools used them in class as guided reading and comprehension exercises and appreciated that they were appropriately differentiated and that they could be used during ‘literacy time’.

“The Year 5s did all of them [extended reading exercises], all three of them. Four of them actually because I did the £10 note one as well. But we didn’t do it as homework, we did it in class when the Year 6s were doing some booster practice SATS. The year 6s were reading practice SATS and they were sitting there doing those reading comprehensions. They were pretty good reading comprehension practice.”

“We used the comprehensions but we didn’t use them in science we used them in English. The Mary Anning one worked well in our mixed class the year 5’s used the lower ones and my year 6’s used the higher ones so that was quite nice to have that differentiated for them as well.”

“As part of literacy for comprehension and guided reading we looked at all three of them. We obviously looked at Darwin quite a bit and Mary Anning, which was really good as a female role model and they really enjoyed that. I used them in guided reading because it's quite difficult to get them into the science lesson and they're doing extra science.”

“We did the Mary Anning one. As they were differentiated, it gave them a bit of choice. So even though I had my low set, I said that this is the more challenging one if you’d like to try that because a few of them do like to do that”

“They did enjoy the comprehensions as well. Mary Anning features quite a lot through the school, so actually it was quite nice to be able to look at her in a little bit more depth. Then it's also led into some nice story writing as well.”

Other schools used the exercises as homework saving them having to plan what to set for that week.

“It was very useful having them on sheets because I don’t often like sending books home because you never get them back. It makes marking very difficult when books go home and I don’t see them for the rest of the week. The homework was very helpful because I’m meant to set one a week, sometimes you do feel a bit like “Ah, what can I set?”, so it’s slotted in nicely.”

It was evident from these comments that the teachers and students particularly liked the Mary Anning exercise for its local relevance and as a female role model.

6.4.6 The practical nature of the resources was widely appreciated

Teachers reported that their students enjoyed the practical nature of the resources which made the topic more interesting and accessible through the use of a variety of interactive learning styles.

"It [the Scheme of Work] was very diverse, and they've enjoyed it. They've enjoyed the practical elements the most, they really liked the surveys with the dimples and the tongue rolling. The traits, because actually that is something they can all work out there and then, and do a survey in the class and make a bar chart. So, it made that really simple idea into something that they could really expand on and start to understand, because it was meaningful to them. Because actually, if they hadn't thought about it before, now they know, A, can they roll their tongue, B, if they can, why?"

"The practical elements of this [Scheme of Work], I think it made it more interesting. I think it makes it enjoyable for them. They do love all that sort of thing. I'd definitely keep that in because they did like that."

"It was really good at targeting all sorts of learning styles as children do respond to being able to move around more and that sort of thing. It was much more interactive for us."

"I think one of the things as a whole school, that we need to be really aware of, is making sure that they have that practical opportunity as it's lovely because they do enjoy it. But and I think some units of work lend themselves to that far easier than others like light. Whereas with this sort of thing [evolution] actually it's a bit harder so it's nice to have those physical get up and move about kind of activities as opposed to just let's look at this picture and then, yeah, write about it."

"Yeah, we loved the scheme, really loved the scheme, loved it. Particularly liked all the practical stuff and the fact that it was all prepared for us was an absolute Godsend. I think it made a difficult subject really accessible for the children, and I would definitely say, as I say they were brilliant."

They also commented that practical activities were more memorable for their students rather than using work sheets.

"I think because anything practical becomes significant to the memory doesn't it?"

"It's having good resources which have got some practical elements, some drawings, some interesting things, games and animations. They loved the topic with a variety of different things to do rather than just reading a worksheet. It's those kinds of things that stick in their mind isn't it? It just made it a bit more interesting than the usual, here's another worksheet which is terribly dull."

"They enjoyed that [the pentadactyl limb]. They thought about the colouring and then onto how they were different. The fact that it was something practical and tactile they really tried hard and they were saying to each other, "Oh yours is brilliant, you've got all those bones in there." I think there are so few opportunities now for practical stuff, so I would say it was a really successful scheme. Really good."

6.4.7 Enjoyable lessons led to increased student motivation

Some teachers reported that the students enjoyed the topic so much that it inspired them to catch up on missing lessons, find out more about the concepts covered and even share their new knowledge with their parents.

"We have really enjoyed it [the Scheme of Work], and funnily enough, it's one of the reasons that we sort of almost didn't have an ending, because we headed off into growing crystals and minerals and things like that. They'd brought in lots of things, so actually, one thing led to another. I think it's important to give them the freedom to go off into something they're interested in."

"Then we did a bit of extra work on that as well, didn't we? To do some... a bit of extra study on mass extinction. From the timeline they found out that it just wasn't just the dinosaurs. They were really surprised about that as well. So, they wanted to research who became extinct... and the reasons for the extinctions. It just clarified their understanding of the mass extinctions"

“We worked on the reading comprehensions that you set, a lot of questions came out of that, a few of them did extra research because of it, wanting to find out more, especially about Mary Anning. They were very interested, so we had lots of extra discussions.”

“They really enjoyed the Trilobites, in fact, some of them were away because they were rehearsing for the school play and missed a bit and they were like, “Oh, but can we still do it, Miss?”.”

“I know a few of them actually went home and told their parents about it [hunting moths activity].”

“We have a seriously dyslexic boy, so his written work wouldn't have been as good. But he went home and researched quite a lot and came in and told me all the things he's found out.”

“The thing we've kept coming back to is the original YouTube clip, because they're just amazed. One of the fathers had quite an argument with his daughter about whales never had back legs, so she went home, showed him the clip and he had to admit that he was wrong.”

These comments support the assertion of Chi (2009) that when children are interested and actively engaged in lessons they are more likely to learn.

6.4.8 There was widespread endorsement of the teaching materials

In addition to the 13 schools who endorsed the project’s resources by repeating their participation in the study for a second year, other teachers reported that they would also be willing to use the course materials again.

“I think when we do it again, no I'm sure I want to use it all again, I think it's going to be even better, now that I know how the lessons go and how they flow and fit together. I think it's going to get even better than this time.”

“It’s been a very positive experience for me. I mean I’d love to teach it again. I must say as well, the power points and the teaching resources for me to look at were brilliant, were really clear and…good to go. And it should be slightly more user-friendly next time because

I know what I'm doing. Teaching it [evolution] has not been a problem at all because I think there was so much to work with, it was easy to pull out and take what was required."

"I wouldn't change anything. They all really enjoyed it actually, they found it fascinating, because it is such a fascinating subject. It was a success. I would definitely use them again, I'd probably use the whole thing."

"This year I was very conscious of your study and tried to follow your scheme as closely as possible. Having done it once and feeling very comfortable now with what is in, I think if I redid it, I would do things perhaps a little bit more differently. I would modify the worksheets and the graphs, and there would be other teaching to tie in with the graph work."

Even the some of the more critical schools who expressed their reticence in following the prescribed Scheme of Work provided by the study felt that they would use certain activities again as part of their own curriculum planning.

"The hand span activity was quite a good exercise but we'd probably use that again in a slightly adapted way next year. And actually, the time line was one that I actually did find that was quite useful, because I was able to use the hall and because it was quite visual the kids could sort of get the idea of a common ancestor."

6.5 Wider Implications of the study

The results of this study have wider implications for the teacher training and how to support non-specialist teachers to enable them to teach evolution confidently in a conceptually accurate manner as well as in the ways we can evaluate the effectiveness of teaching interventions.

6.5.1 Issues with religion did not impact on teaching

During the course of this study a small minority of the teachers (predominantly in church schools) experienced some minor issues with religion, ranging from questions about how evolution fits into biblical accounts of creation to a student being withdrawn from lessons. The resources developed by this project were not designed to conflict with or intentionally force students to change their own personal religious beliefs. It was intended that an advocacy approach would be used by the teachers in order to introduce evolution. This approach, one of three different instructional approaches identified by Reiss (1992) promotes the use of a Scheme of Work that moves student conception of evolution to be more in line with the accepted scientific understanding of the topic without discussing or expanding on their alternative conceptions, thereby facilitating conceptual change by exposing faults and promoting dissatisfaction in their beliefs. In practice it is incredibly difficult to teach evolution using an instructional approach without questions about religion arising. If such questions arose they are only discussed to expose and correct alternative conceptions removing the controversial nature of evolution.

"There were a few comments because… In the Year 6s a lot of them are in the worship group and plan assemblies and things, but some of them did make a few queries about "Well, if we've all come from this one billion years ago, I thought we came from Adam and Eve." But no, it wasn't… It was just a few little comments, not anything big. Yes, even though we're a religious school, a Christian school, it wasn't a problem. It was more an interest than it was a "Oh, that can't work if I believe this." Yeah, it was trying to get it in their heads really. We didn't have any huge discussions about it. They just made a few comments".

"There were a couple of questions. We don't have any Creationists as far as I'm aware but we are a church school, so there were a couple of questions comparing the version in the Bible to this, but nobody was so sort of hard and fast, no the earth was created in seven days, so this cannot be right. So, we just talked about fossil evidence being the proof, scientific evidence. Whether or not I've answered their questions satisfactorily or not, I don't know. We've come to these conclusions because of the evidence, yeah. I mean none of the questions were sort of challenging to scientific evidence, and I've got a couple of little atheists who were quite you know, "That's rubbish, that's just a story." So, it was a balanced sort of thing".

Several teachers endorsed the compatibility between the acceptance of evolution and a belief in God, and used coexistence models in their classrooms similar to those reported in college educated adults Brem et al. (2003) and adult museum visitors Evans et al. (2010).

"I'm a Christian and in some of the lessons, particularly towards the end when, inevitably, it's how did the world come into being and how did these things happen? And it, especially the timeline threw up lots of questions about how did that happen. And you've got to go, "Well people believe different things," and it did kind of open up bigger questions, and I think it's quite a hard, it's personal to people and to their background. So, people from different religions would have different stories, some people would think it that God created the world and others don't".

"We had quite a lot of questions because we're a church school, so we discussed that as well. We did talk about the fact that this is a theory of evolution just other science ideas, you know, is a theory based on evidence and that we also, we've got people in school who are quite religious. We explained to them that although you can believe in God, you can also believe in evolution, so they can go alongside each other, which I think was great."

"We did sort of mention that evolution can be a controversial subject to the children, I mentioned that, you know, that this is the theory of evolution and there are some people who don't believe in this, but this is a scientific point of view that lots of people believe in, you know, and a lot of very religious people, believe this as well, so you can still believe in evolution and be religious".

The majority of teachers did not report any problems regarding religious beliefs in their classes even with religious students.

"Definitely, they really enjoyed the topic, and the subject of it. It was surprising actually, because I think when we met before I asked if you'd ever had any difficulties with children who were quite religious. It was strange, because the boy who I was thinking about in particular when I asked that question, he enjoyed this topic the most, and he was really in to it, so there were obviously no difficulties".

Only the beliefs of Jehovah's Witnesses caused any religious conflict with the teaching of the evolution topic in three of the schools in this study, as they are based solely on Biblical text. They believe that the Bible is the actual Word of God and consider its 66 books to be divinely inspired and historically accurate. One of the teachers pre-empted any potential problems by discussing the Scheme of Work with the parents before teaching the topic.

"I had just an informal chat about a student and her Mum brought it up that they were Jehovah's Witnesses. I invited her in, because I wanted to check…I went through the resources with her first, basically just used your folder and said "This is the plan," to see what she would be happy with. And she was very supportive. Yeah, she was fine. She just said that "I have different reasons for them sharing the same bone structure" and things like that, and that's good because we don't want kids missing science. And she was actually one of them that particularly enjoyed this activity."

Another teacher reported a religious student's dissatisfaction with the assessment instrument feeling that there wasn't an opportunity to disagree with the response options offered for each item.

"We have a little boy in here who is Jehovah's Witness. I'd already delivered the unit of work, and the father probably would have liked it if he'd known about teaching evolution beforehand. It was the questionnaire at the end post-study, with the multiple-choice questions when I think the boy in question felt he had to answer. What I think he would have chosen to do was say, "I don't agree with any of these statements," but he didn't feel that he could, and this was the point that the father was making. He obviously talked about it at home and perhaps felt, I don't know, awkward about it. He could have chosen to opt-out, it doesn't happen that often, he won't attend certain assemblies and at Christmas time he doesn't get involved in making all the decorations and Christmas cards and stuff…but we're teaching that this as science with facts."

Only one student in the entire study was actually withdrawn from participating in the study by his parents.

"*I had a Jehovah's Witness student who was withdrawn from class during the topic. His parents weren't happy for him to sit in, but obviously he missed out"*

6.5.2 The project provided support for non-specialist science teachers

Many of the teachers in the study were either non-science specialists or teaching the subject for the first time. From the training and support given by this study they felt they were able to teach the topic successfully using the Scheme of Work and resources provided by the project.

"I'm a complete non-specialist. Yeah, it [the Scheme of Work] was fine because I just had to read it first and be up to date on it before I delivered it but it was clear enough for me to be able to do that."

"I think myself and one of the teachers have got science degrees, but everybody else has not got a science background at all, and in fact the lady next door to us, she's never taught Science at all here or anywhere else before. So, she said she quite enjoyed it because I

usually take her class for Science, so it was nice for her to be able to do something like that."

"Obviously as a Newly Qualified Teacher (NQT) this is the first time teaching this subject. I haven't had any experience, not really, apart from just what I'd done at school myself, but I only did science up until GCSE, and then things you see on TV and things. I would hate to teach them something that wasn't accurate."

"With all the resources and everything in here, it was fantastic, and the teacher's notes really helped prepare me to teach it "

"If I hadn't had the resources I wouldn't have taught it in the same way. Adaptation is probably fine but I wouldn't have had the simulations and I probably would never have come across or been brave enough to attempt the limbs. If I just glimpsed it on TES, I would have thought, "Oh, no, they can't do that." Or "Oh, I'm not sure what's going on there. If you just see the resource you're thinking "What am I meant to be doing?" but with a combination lesson plan and their worksheet with the objectives and the instructions on, then you can't really go wrong."

Several of the teachers were concerned about the relevance, accuracy and utility of the other resources available to them to teach evolution. They also were concerned about the possibility of intentionally teaching their students misconceptions from their own choice of resources.

"You can go into the TES, where there's lots of things but you don't know if they're relevant if they're accurate or if they work"

"I can honestly, it [the project] meant that we've delivered this in a far, far better way than we would have done if it was down to us because we, with our science, at the moment, kind of scrabbling around for schemes and…as I say, we would have ended up hunting around to try and find something that was out there…"

"Especially I think with this topic, you need to be mindful that you're teaching it correctly and you're not giving them any misconceptions. To be perfectly honest, I probably would have gone online because I would want to make sure what I was teaching them was correct rather than me fumbling around…But this would have been very time consuming to go and

collate all of those different resources from different areas to use, whereas this was just really succinct, and I followed it."

They also acknowledged that the Scheme of Work and resources provided by the project had saved them a lot of preparation and planning time allowing them to teach the topic in a more imaginative and flexible way.

"I think it's much more accessible and fun for the children than some of the other planning that I've looked at."

"You can see how under resourced schools are for science by the fact that when we get out these little bags with the moths and forceps in the children are, "Wow". For then to be so excited over that, just shows that imbalance of how often they get things and the funding and everything. So, I think that's been really important to recognise, and how key it is to get this done. Really, if we hadn't have had the resources I don't think we could have taught the lessons, obviously just because teachers have so much on their plates, and it's just the ideas, knowing what to do and the prep. So, thank you for all of that, thank you very much."

"You've given us a basis for the lessons, you've given us the resources and the ideas which I would have spent hours trying to find because I had started to look before we met. And because it would have taken me so much time and it wouldn't have been as imaginative."

"I really enjoyed teaching it [the Scheme of Work]. It was easy and I didn't have to put in much time in the lessons. It was pretty much all there, good to go. Because I mean with all the best will in the world, you don't get to read a book. You can't prepare everything can you in this much detail. There's just not enough time.

"I really enjoyed it and it was a program that I knew exactly what I was going to do for so many weeks. That for me was perfect. It's been great to be honest. Perfect not to have to plan for a topic."

"I liked the way that we could do the plenary first and then go back to the main parts. There was no real, "This is how it should be done." It meant that we could fit it in much better, because some of it broke down into smaller slots of 20 minutes, others were hours, others were two hours and we might have done a whole afternoon. So that worked well."

They also commented on their difficulty in translating the National Curriculum criteria for the topic to appropriate, accessible activities within a coherent framework, in line with the teacher case studies of (Crawford, 1999; Marx et al., 1994).

“For me as a year six teacher in the new curriculum I found it really difficult at the beginning to know where to start with such a big topic of evolution and inheritance. The national curriculum doesn't give very clear or detailed guidance, although it is quite restrictive in terms of what you have to teach in each year group. So, I found it quite difficult, more in the early stage, knowing where to start. Certainly, last year my whole approach was more research based, there weren't many resources, and certainly not resources that weren't very expensive available. I didn't have a good enough subject knowledge to know really how to turn that into a really practical based set of activities.”

6.5.3 Participation in the study led to increased teacher confidence

Teachers participating in this study reported that they felt supported by the training and resources provided which lead to improved confidence in their ability to teach the topic, confirming the results of the quantitative analysis (see section 5.3.4).

“I thought they [the resources] were fantastic and really supportive for me with an area of the curriculum that I didn't feel that confident with and they were really interactive and engaging for the children. They thoroughly enjoyed it and I think they got a lot out of it.”

“I think this year, having the project has given me lots of ideas and lots of hooks in, and then we could then expand on that, or I could see how that linked in, and just helped me have a bit more confidence with it really. I'm certainly more confident actually having the resources provided, having the time and the guidance just meant I could be really clear, good resources online as well that clearly stated videos in particular.”

“I really enjoyed it, I thought the resources and having them already made for me a lot easier. Then it gives you a lot more confidence to teach it knowing that it's all there ready for you. Then I liked having the background information, so I read that before teaching each lesson so that I was definitely clear on what I was teaching, and I wasn't going to be saying anything that wasn't quite right.”

“Having done it already with it [the Scheme of Work], I would feel more confident about doing it next time. It's been great for me.”

Teachers participating for a second year commented on further gains in confidence leading to improved effectiveness in the classroom.

“I think the first time I taught it, I was fairly confident and I’d say I’m more confident now I’ve done it twice. I’m not really confident yet, but I’m definitely more confident.”

“I think I was much more confident with the content and I get the impression that I actually led it much better this year. But I think that's down to me maybe doing the second time round as well. I think I just felt more confident with it.”

6.5.4 Being part of study acted as a motivator for students

Some teachers reported that their participation in this study was a motivating factor for their students who liked the thought of their results forming an important part of a larger science project.

“That was my concern, that they might muck around, they didn’t, they were absolutely brilliant and they were really serious and they really took it seriously. I said, “You know, this is a really important project, we’ve been chosen,” and they absolutely loved that whole thing, that they were working for you. I put something in the newsletter about it too.”

“I think, they felt quite privileged that they were being used as guinea-pigs. ’Cause we kept saying, “This is for the university, you’re part of a research project” And they sort of rose to that. They enjoyed that. Definitely, a really positive experience.”

“On the whole they did really enjoy taking part and actually they quite liked knowing it was for somebody else because I'd explained that we're going to be doing some of our learning for you. It was a nice change I think, rather than it always just being for me in their eyes.”

6.5.5 The study relied on teachers to adhere to the scheme of work

Classroom observations and comments from the teachers demonstrated that they were very conscientious in following the Scheme of Work as closely as possible. However, some teachers expressed difficulty in following the prescribed scheme of work as they would have preferred to do their own ‘thing’. This is problem for this research as it needed the teaching of the topic to be as standardised as possible, as the teaching and learning styles employed in different classrooms for the same topic can vary dramatically depending on the personal preferences of the teacher and the ability of their students.

"For me, reflecting on it, it's that kind of supply teacher syndrome, I find it really difficult to work from somebody else's plans. I think we all do as teachers, don't we? I think ooh I don't know, would I have done it that like? Knowing my subject knowledge of evolution is, is to say the least lacking, the resources were brilliant, obviously, from that point-of-view, but it was just, well, I got a bit, frustrated but I gained confidence as I taught it."

"It was all organised. It was all there. It took me a couple of lessons to just get into the swing of the style because we all have our own way of doing it, don't we? But yeah, I thought the lessons were fine. Some of the activities I struggled to get my head around initially but actually once I got going it wasn't so bad."

6.5.6 Teacher understanding improved by participating in the study

Participating teachers also commented that they had improved their own understanding through teaching the topic. This is evidence of metacognitive development with teachers gaining a greater understanding of their own learning process, becoming more aware of the partial nature of their understanding.

"I would say that I think it's refreshed my knowledge quite a bit."

"I've definitely learned some stuff teaching the topic. The timeline. I was a bit amazed by that actually and I would have definitely got some of those times wrong or the dates wrong. And definitely with the limb. That wasn't clear to me before doing this. It's something I'd not really thought about as evidence. You just say fossil evidence, fossil evidence. That's the term that gets banded around but you don't actually pick that apart."

"Rather embarrassingly, I found it extremely difficult filling my questionnaire about the guppies and the lizards [CINS] before I taught it. Because I was just looking at the statements and thinking, 'It could be that one,' and then, 'No, it's probably that one.' I think I could maybe do better now if I did it again."

6.6 Chapter summary

In this chapter qualitative feedback obtained from participating teachers has been summarised. Teacher opinions on the utility of the different resources provided by the project and how they improved the learning of students in their classes have been explored enriching the findings of the quantitative analyses, as well as considering the benefits of participation in research projects to both students and teachers.

Chapter 7 Discussion

7.1 Chapter overview

This chapter brings together the quantitative results from the two large scale randomised trial tests and the qualitative data obtained from participating teachers in order to draw conclusions on the overall effectiveness of the teaching resources and training provided for primary school evolution education. It also considers some of the limitations of the study and interesting points that arose from conducting this research that may have potential implications in the wider field of evolution education.

7.2 The Big Picture

7.2.1 The first large scale repeated randomised trial tests in primary evolution education

The introduction of the Inheritance and Evolution topic into the Key Stage 2 National Curriculum for the first time in 2014 provided this study a unique opportunity to research evolution education in UK primary schools. Very little is known about the understanding of natural selection and evolution in primary age children with around half of the existing research with this age group relating to the understanding of inheritance and kinship (Kargbo et al., 1980; Solomon, 2002; Solomon & Johnson, 2000; Springer, 1992, 1995, 1996; Venville & Donovan, 2005). Although more recent evolution education research has focused on exploring the understanding of natural selection and evolution (Berti et al., 2010; Browning & Hohenstein, 2013; Chanet & Lusignan, 2008; Kelemen et al., 2014; Venville & Donovan, 2007) these are mainly small scale experimental studies or are editorials based on opinion written without substantive evidence (Campos & Sá-Pinto, 2013; Wagler, 2010). The majority of data collected from structured oral interviews or analysis of lesson products, there being little quantitative data available on effective teaching approaches as supported by Beardsley et al. (2012). (see Appendix A for a full summary of existing research with primary age students).

This study is the first to present repeated results from a randomised trial test taken from two large-scale tranches of data (tranche 1 n = 1152 students collected from 17 schools, tranche 2 n = 1505 students collected from 28 schools) into the effectiveness of different evolution teaching intervention programmes in 9 to 11 year-old students in the UK. The study compares the relative effectiveness of four treatment groups with each other based on student a priori knowledge of the topic. Participating schools were randomly assigned one of four Schemes of Work (Scheme of Work) after recruitment to the project and then received standardised resource packages and teacher training. The four groups of schools were followed in exactly the same way the only difference between them being the Schemes of Work taught in class. Care was taken to randomise the allocation in order to minimise bias, balancing both known (school type, size, location) and unknown predictive factors.

7.2.2 The assessment instrument was fit for purpose

The results of this study have provided evidence that the student assessment instrument and its mode of delivery were appropriate for use with 9 to 11 year-old students in the UK. The level of difficulty allowed clear discrimination between students of all abilities, with modification of the 'readability' of the assessment items enabling students to access and understand the difference between alternative responses. The novel mode of delivery was quick to complete mitigating both poor literacy skills and some of the negative effects of question fatigue. We have demonstrated reliability of the assessment instrument by the consistency of the results across the two tranches, demonstrated by the strong test-retest correlations between the percentage of correct answers per item in the pre-test ($r = 0.98$, Pearson's correlation coefficient) with no significant difference between the LOESS residual change in score ($P = 0.26$, Wilcoxon rank sum test) and across items, with acceptable and improving internal consistency scores (Cronbach's coefficient alpha) over successive tests. The assessment instrument was also valid with supporting criterion related evidence from the correlation between student scores and their independently assessed teacher judgement of their science ability levels, with strong ecological validity as conducted in their own classrooms and finally construct related evidence from the analysis of the data itself.

7.2.3 Students made significant gains in understanding

The evidence presented by this project supports the growing body of research (Evans, 2008; Kelemen & Rosset, 2009; Legare et al., 2013) which suggests that upper primary school children (aged 9 to 11) possess the cognitive ability to successfully understand the concepts of natural selection and evolution, when provided with appropriate resources and teaching instruction, with no significant difference in performance due to student age even in mixed age classes found in smaller primary schools. It also supports the premise that primary students can successfully learn about natural selection (Metz, 2010; Nadelson et al., 2009) and understand the basics of the principle if appropriate scaffolding is provided (Evans et al., 2015; Metz, 2010).

Evidence was consistent across both tranches showing significant improvement in student performance in both the post-teaching and retention tests in all four Schemes of Work (Scheme of Work). The magnitude of the effect sizes above the 0.4 ‘hinge point’ for useful teaching interventions as defined by Hattie (2012). However, the degree improvement varied from item to item with time, depending on the resilience of the alternative conception in question, some being more resistant to modification than others:

1. This project succeeded in correcting essentialist thinking (Emmons & Kelemen, 2015; Gelman & Rhodes, 2012) moving students away from the concept of species ‘essence’ and enabling them to appreciate the importance of intraspecific variation in order to understand the process of natural selection.

2. Instruction also proved useful in helping students understand that members of different species can share a common ancestor even if they don’t necessarily share a lot of similarities. However, students still found the human and chimpanzee pairing in assessment item 1 to be particularly resistant to conceptual change. In the future it would be interesting to assess whether two other closely related species would give similar results if substituted e.g. dogs and wolves, or whether this was a human-centred phenomenon.

3. Anthropomorphic alternative conceptions, those involving soft inheritance and Lamarckian thinking also persisted after instruction.

4. The instruction provided by teachers participating this project reduced the frequency of all types of alternative conception and so was successful in forming the first ‘rung’ on which to build upon in the spiral curriculum for evolution education (Venville et al., 2005).

Importantly, students of all abilities (as judged by their teachers) responded positively to the teaching interventions making significant improvement in their matched raw pre-post scores e.g. low ability students ($P = < 2.2x10^{-16}$, Wilcoxon signed rank test, tranches 1 and 2), suggesting that resources developed by this project were appropriate for all abilities in mainstream schools.

Student gender was found to be a weak predictor of student performance with girls significantly outperforming boys in tranche 1 ($P = 0.01$, Wilcoxon rank sum test) but with negligible effect sizes in both tranches of data (Cliff’s d = 0.09 tranche1; 0.06 tranche 2). This result is in line with large meta-analyses of gender effects on examination performance (Hyde & Linn, 2006; Wilkinson, 1999) who found undue emphasis placed upon statistical significance while effect sizes were ignored exaggerating the magnitude of gender differences. It also supports the conclusion of the National Pupil Database Key Stage 2 findings which stated negligible differences in student performance at KS2 in 2017.

(https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/667372/SFR69_2017_text.pdf)

7.2.4 Main activities interact in reciprocal and reinforcing manner

Schemes of Work 2 and 3 were found to be more effective relative to the other two programmes of instruction due to the unforeseen reciprocal interaction between pairs of main activities in lessons 2 and 4, exposed by the large sample sizes we employed. The teacher-centred Peppered moth power point activity was more effective when taught in combination with studying homology in Trilobites, whilst conversely the more student-centred 'hunting' moths activity was most effective when combined with studying homology in the pentadactyl limb. The exact nature of this reciprocal relationship is intriguing and warrants further exploration in the context of evaluating sequences of activities within teaching intervention programmes, rather than stand-alone single activities.

Whilst the cause of this interaction cannot be identified the highly repeatable design of this study allows a number of factors to be controlled and excluded. In both tranches Schemes of Work 2 and 3 were the most effective, followed by 4 with 1 being the least effective of the four teaching intervention packages. This order effect is unlikely to be caused by the uneven distribution of 'better' teachers across the schemes as Scheme of Work 1 did not have a significantly lower distribution of teachers with high confidence scores, years of experience, understanding or acceptance of evolution.

The results suggest that the main activity in lesson 2 acts as a primer for the main activity in lesson 4. As we used the Peppered moth to show the process of natural selection in lesson 2 in both alternative activities the difference cannot be due to the context of the activity but rather the nature of the activity carried out (student or teacher-centric). Conversely, in lesson 4 the same activity was employed to investigate common ancestry through homology but this time in two different contexts (extinct Trilobite species and extant mammalian pentadactyl limbs) and so the difference cannot be due to the nature of the activity but caused by the to the context of the activity. This sort of interaction is seldom considered in educational studies and will be the focus of a future research paper.

7.2.5 Teacher confidence levels and experience are important

Importantly, no evidence was found that participating teachers were significantly hindered in their ability to teach natural selection and evolution effectively using the resources supplied by their own understanding and acceptance of evolution, religiousness, highest biology qualification, formal evolution education, gender or the number of times they used the resources. Only three factors had any explanatory power; teacher perception of their increase in confidence after teaching the topic (both tranches), completion bias (tranche 2) together with their years of experience (tranche 2), however, none remained significant after multi test correction. The reasons for these results are purely speculative and could include higher levels of motivation and enthusiasm in teachers reporting larger gains in confidence and the opposite effect in 'co-opted' teachers. Although years of experience do not necessarily lead to teaching expertise (Berlinger, 1987), the results of tranche 2 suggests that more experienced teachers were more effective, perhaps due to their superior pedagogical knowledge (Wilson et al., 1987) and pedagogical content knowledge (Sanders et al., 1993). We propose that participation in the project acted as professional training for participating teachers allowing them to increase their knowledge, confidence and understanding about teaching evolution, enabling them to teach the topic more effectively, this premise being supported by the qualitative feedback (see sections 6.5.3 and 6.5.6). This finding is also in line with the conclusions drawn by the National Research Council (2017) report regarding the importance of continued professional development programmes for teachers. (https://www.nrc.no/globalassets/pdf/reports/nrc-global-education-strategy/nrc---global-education-strategy-2018-2020.pdf)

7.2.6 Primary schools performed better than Middle schools

Of the 14 school level factors investigated only the type of school attended was a significant confounding factor in tranche 2, with a combined $P = 0.03$ (Fisher's test) and a large effect size (Cliff's $d = 0.73$) with students in primary schools performing better than those in middle schools. Middle schools might have been predicted to perform better than primary schools as they have a higher proportion of specialist science teachers, however, it seems that this was not an important factor. The results suggest that primary school teachers were more effective overall and this could be due to a number of factors relating to structural differences within the schools or better pedagogical practises in primary classrooms. For example primary teachers may 'know' their students' better as suggested by Brühwiler and Blatchford (2011), have greater lesson flexibility as they are not confined to rigorous subject timetables, place greater emphasis on cross curricular teaching and autonomous learning (Anderson et al., 1979) and perhaps have higher levels of enthusiasm for the use of the resources provided. Additional explanations can be found in the research of Eccles et al. (1993) which compared maths teaching in primary and middle schools and found that middle school teachers tended to control their students more and provided fewer opportunities for discussion. These factors were not considered in the design of this study but would be interesting to follow up in future work. It should be noted that school type as a of predictor of student performance was not significant after multi test correction. Additionally, when the student level analyses were repeated after removing middle schools and combining the smaller remaining data sets (n = 1249) very similar LOESS residual results were obtained, suggesting type of school was not a significant predictor of student performance.

7.2.7 Teachers held similar alternative conceptions to their students.

Analysis of the results suggest that while teachers in this study showed a high acceptance of evolution they did not have an adequate understanding of the process of natural selection. This lack of understanding mainly resulting from alternative conceptions relating to anthropomorphic thinking and soft inheritance mirroring those reported in their students post instruction. This could demonstrate the persistence of these deeply engrained alternative conceptions into adulthood (Bloom & Weisberg, 2007; Nadelson & Sinatra, 2008) or it may suggest that the initial teacher training provided pre-instruction did not adequately addressed these alternative conceptions.

7.3 Limitations

7.3.1 Representative sampling

The representativeness of this study sample is limited to 9 to 11 year-old students and their teachers in the Southwest of England. Participants being predominantly white of European or British decent. As the generalisability of any study is limited to the population from which the data is obtained (Mpeta, 2014; Van der Riet & Durrheim, 2006), other groups within the UK population might have yielded different outcomes.

As teachers had to volunteer to participate in the study the sample was probably not a truly random sample (Yates & Marek, 2014). The self-selecting nature of recruitment may have produced a sample biased towards well motivated classroom practitioners. However, this bias was not of concern as the study was not attempting to make generalisations about the whole teaching population (Cid, 2013). Instead it was an exploratory study examining the effects of different Schemes of Work delivered by teachers who were highly motivated to participate evident from the time commitment they were willing to dedicate to the teaching of this topic in an already packed curriculum.

7.3.2 Fixed lesson order

Our aim was to provide an evidential basis for establishing best practice. In this regard we note that due to the cognitively progressive nature of the concept the order of the 4 lessons in each Scheme of Work could not be altered (variation -> natural selection, -> geological time -> evolution). Therefore, the relationship between lesson order and understanding was not studied. This best practice restriction accords with a large-scale randomized trial indicating that, at least in secondary schools, teaching genetics before teaching evolution causes a marked improvement effect on evolution understanding, at no detriment to genetics understanding, compared with the opposite treatment (Mead et al., 2017)

7.3.3 Unintentional formation of a new alternative conception

From the analysis of alternative conceptions detailed in chapter 5 it appears that teaching the topic using resources provided by the project unintentionally instigated the formation of a new alternative conception; a post instructional increase of 7.83% in the frequency of the incorrect response - that extinctions are only confined to rare mass extinction events (see assessment item 11). This alternative conception may have resulted from the emphasis placed on the five main mass extinction events in the time line activity in lesson 3. To overcome this problem perhaps an additional comment could be included on the terminal teacher summary card to explain that extinctions have occurred throughout the history of the Earth and are not merely confined to mass extinction events.

7.3.4 Reliance on teachers to adhere to the scheme of work

Due to the large sample size and number of classes participating in this study it would have been impossible for the principle researcher to be responsible for the teaching of the topic due to time restrictions. For this reason, instruction was carried out by the designated classroom teacher using the project's resources and following the allocated Scheme of Work after detailed training was provided, also giving the study strong ecological validity. Teaching of the project was therefore completely dependent on the teachers delivering the lessons in a standardised way, the fidelity of the results relying on their experience and professionalism to 'teach to the scheme'. Although a small minority of teachers did not complete the salt dough modelling component of lesson 4 to consolidate the homology colouring activity in their classes (6.80%). Data from these classes was still collated and analysed en masse as these omissions occurred evenly across all 4 teaching intervention programmes and all other learning experiences within the lesson were completed as evidenced by classroom observations and teacher feedback.

7.3.5 Improvement in teacher understanding could not quantified

Although participating teachers commented in the qualitative data that they had improved their own understanding through teaching the topic (see section 6.5.6), we were not able to quantify these perceived improvements as only one full teacher questionnaire was completed before teaching the topic which was followed up by a shorter survey to assess any confidence changes. This was a planned strategy employed to minimise ‘respondent burden’, defined by Sharp and Frankel (1983) as the time and effort involved in taking part in a survey. From this feedback it is evident that they perceived their understanding of the topic was improved by actually using the resources and teaching their classes supporting the research of Schank (1995).

7.3.6 Possible evidence of survey fatigue in the retention tests

No significant evidence of question fatigue in the pre or post-test sessions in either tranche was found when the percentage of non-response (NA) or ambiguous (U) answers were correlated with increasing assessment item number. However, significant correlations with moderate effect sizes were found in both retention test results for both of these factors demonstrating clear evidence of question fatigue in both tranches, although the percentages in question were very small (NA mean % = 0.23 ± 0.25; U mean % 0.31± 0.2). One hypothesis could be that these results may be caused by survey fatigue in which multiple surveys appear to suppress response rates rather than question fatigue. What is not clear from these results is why the retention tests which were repeated ~ 18 weeks after the post-tests showed significant fatigue rather than the post-test carried out much closer to the pre-test session as expected from other research (Porter et al., 2004). One possible explanation could involve the timing of the final retention test within the school year, the majority of the these tests taking place near the end of term 6. This is just after year 6 students have to complete a rigorous national programme of Key Stage 2 Standard Assessment Tests (SATs) which could have resulted in lower student motivation levels due to ‘over-testing’. Therefore, the scheduling and number of individual tests within the school calendar needs careful consideration during the planning of any intervention programme.

7.3.7 Potentially biased retention sample

The retention sample may have been biased towards completion by more highly motivated teachers or those who had built up a stronger personal relationship with the principle researcher. Another key factor in the completion of the retention tests and was the scheduling of the topic within the school year in the individual schools. This decision was made on a school to school basis with only those teaching the topic early in the year (before Christmas) being able to fit it in before the end of the academic year.

7.3.8 The student questionnaires were perceived to be too hard

Genuine concerns were raised by participating teachers in their qualitative feedback (see section 6.3.1) regarding the appropriateness of the assessment instrument developed for use in this study. The majority of teachers thought that the questionnaire was too difficult and worried that the assessment would not reflect the improvement in understanding they observed in the classroom. However, the results presented by this study demonstrate that this assessment instrument was both valid and reliable, allowing improvements in student understanding to be measured. The level of difficulty of this assessment instrument was carefully planned and trialled in pilot schools to allow the ability to clearly discriminate between individual student performance as well as between option choices within separate assessment items. The results do not support the premise that the students were merely guessing, giving random answers to the assessment items because they were unable to understand the questions. Instead, a strong agreement in the types of alternative conceptions chosen in the pre-teaching questionnaires was found across the whole cohort which then altered significantly post instruction. These homogeneous and significant changes in student understanding suggest that the student questionnaire was a suitable assessment instrument for this research project. However, if the teacher feedback is considered it is likely that the gains in understanding measured using this assessment instrument whilst being significant, represent the lower bounds of improvement and so are likely to underestimate the actual learning that took place in the classroom.

7.3.9 Further acquisition of learning cannot be ruled out

Due to the time delay between the post and retention the possibility that students may have continued to learn and improve their understanding of the topic cannot be ruled out. Whilst none of the schools carried out any extra formal instruction on the topic during the intervening time, further reference to and discussion of the topic could have occurred, which could have affected student understanding in the retention tests.

7.3.10 Possibility of positive testing effect

Pre-testing and post-testing using the same assessment instrument is a commonly used method to assess student learning gains (Bao, 2006), however, some studies have shown evidence of positive testing effect, i.e. taking a test leads to an improvement in learning compared with studying alone. Memory research and educational psychology experiments have repeatedly demonstrated that taking a test on studied material promotes subsequent learning and retention of that material on a final test (Bartlett, 1977; McDaniel et al., 2007) This type of research has established that positive testing effect is robust and occurs across different types of study materials and test type; occurring with word lists (Hogan & Kintsch, 1971), paired associate lists (Carrier & Pashler, 1992), pictures (Wheeler & Roediger III, 1992) and prose passages (Roediger III & Karpicke, 2006). Crucially, these experiments have involved testing the recognition or recall of specific, targeted pieces of information contained within the study materials rather than the assessment of learning out comes from a programme of instruction. Most were small scale studies carried out with much older students (usually undergraduates) in laboratory settings rather than the classroom where there is greater variability across students and testing time periods. Additionally, the time periods between the initial and final test were relatively short, typically 1-2 days (Carrier & Pashler, 1992; Hogan & Kintsch, 1971), with only a few studying the effects over longer time frames of up to 1 month (Butler & Roediger III, 2007; Wheeler et al., 2003). Despite the large body of experimental memory testing evidence there are only a few educational studies that have investigated the positive effects of testing (Glover, 1989; Spitzer, 1939).

Although the occurrence of positive testing effect cannot be eliminated from any research that measures improvement in student learning by testing, it can be posited that the magnitude of any positive testing effect in this study would be relatively small and unlikely to be significant for several reasons:

1. Unlike the existing studies into positive testing effect in which the multiple-choice questions within the tests mirror information provided in the studied material allowing recognition of key words, the content of each assessment item within this instrument did not feature in any of the teaching intervention materials, thus limiting student exposure to this material to the tests alone. This allowed the quantification of gains in learning outcomes or overall understanding from the sequence of lessons to be measured rather than recall or recognition of target material.

2. Due to the mode of test delivery students only had access to the assessment items on a screen for a relatively short time minimising exposure to this material within each testing period.

3. No feedback was provided during the study as to the correct answers so that students were unable to correct errors in their understanding (Butler & Roediger, 2008) or confirm correct responses (Butler et al., 2007). This was an important part of the methodology as feedback has been shown to be particularly important after multiple-choice question tests to help prevent students repeatedly giving the same incorrect answers (Butler & Roediger, 2008). Additionally, participating teachers were not given the correct answers to the student assessment instrument to avoid any unconscious bias when they read out the questions.

4. It is probable that there was some degree of negative suggestion effect, an increased belief in incorrect information that the students acquire through testing (Remmers & Remmers, 1926) as the incorrect answers were common alternative conceptions acting as attractive distractors or lures. Taking tests without feedback has been shown to give a slightly higher proportion of incorrect lure responses on the final multiple-choice question test compared to not taking a test (Butler et al., 2007). As no feedback was given these lures could be believed to be true and given again as research has shown that errors tend to perservate, i.e. initial errors tend to reappear despite receiving feedback (Cunningham & Anderson, 1968; Elley, 1966). Other studies have shown that it usually takes several tests and feedback cycles to overcome this tendency to repeat errors (Kaess & Zeaman, 1960; Roediger III & Karpicke, 2006).

5. Similarly the act of encountering false statements even when students know they are false can make them seems true at a later time. This mere-truth effect would also reduce the positive effects of testing as supported by the research of Kelley and Lindsay (1993) who found that undergraduate students could recall previously selected incorrect multiple-choice question lures more easily, this retrieval ease being misinterpreted as confidence in the correct answer.

6. Finally, the study had a much longer retention period than most other studies allowing time for ‘significant forgetting’ as demonstrated by Larsen et al. (2009) in their study on retention of understanding in medical residents. This longer retention period reducing the magnitude of any positive testing effect further.

7.3.11 Potential teacher bias in their assessment of student ability

There is a widespread assumption amongst educational professionals and researchers that teachers are generally poor judges of the attributes of their students and that their perceptions are often subject to bias and error, as identified by Hoge and Coladarci (1989). Unconscious teacher perception biases have been shown experimentally pertaining to special needs provision; boys being more likely to be referred and receive provision than girls (Bernard & Clarizio, 1981; Green, 1993) and be referred for behavioural rather than academic reasons (Croll & Moses, 2000). Socio-economic biases have also been shown to exist in primary age experimental studies (Elhoweris et al., 2005; Pigott & Cowen, 2000) supported by the meta-analysis of Tenenbaum and Ruck (2007) that showed teachers were less likely to refer African American students for gifted programs compared to white students with a difference of 0.92 SD. Teacher expectations have also been shown to affect the way in which teachers perceive and grade student performance, Sprietsma (2013) finding a small but significant positive effect on the grading of more German sounding names by German teachers.

Participating teachers were asked to rate their students' relative science ability (high, middle or low) in order to check the appropriateness of the resources for all abilities and the validity of the assessment instrument. This norm referenced domain specific judgement was then used as a measure of student science ability in subsequent analyses, but how accurate are these estimates? Whist the teacher judgements of science ability analysed in this study cannot be used as proxy IQ scores, the research of Südkamp et al. (2012) suggests that there should be a good correlation between teacher judgements of ability and actual student test performance. In their meta-analysis of 75 studies over the last 20 years they found a positive and fairly strong correlation ($r = 0.53$) between these two factors. The data analysed for this thesis showed strong positive correlations between pre-test scores and corrected learning gains when stratified for increasing teacher judgement of science ability in both tranches, confirming higher a priori understanding of the topic and greater learning potential in more able students as expected by cognitive theory, supporting the that premise teacher judgements of ability were both valid and reliable for use in this study.

7.3.12 Learning gains cannot be compared to a control group

During the planning phase of this study the viability of conducting a randomised control trial (RCT) was considered; see Connolly et al. (2018). This would have involved measuring the progress of students participating in the project against a control group of equivalent students who were continuing as normal i.e. being taught by their teacher without using any of the project's resources or receiving any training. This would have allowed the assessment of whether the average progress made by students in the intervention group exceeded students in the control group; any difference in progress probably caused by the effects of the intervention. However, to obtain meaningful results from a randomised control trial the intervention and control groups must be equivalent, achieved by randomly allocation of students into similarly sized large groups.

Due to the planned size of the samples that were to be collected (>1000 students per tranche) the recruitment of a similarly sized matched control group would have been extremely hard, if not impossible. This project was very successful in recruiting large numbers of teachers as they gained something from participating in the research; receiving new resources and training on how to use them. Teachers in the control group would have had no such motivating incentive and would therefore have been be less likely to test their students in order to participate in the research. Therefore, whilst the results of this study cannot definitively determine whether the teaching intervention programmes developed by this project were more effective than the way the topic is normally taught, evidence has been provided showing that all four Schemes of Work were successful in significantly improving student understanding of the Inheritance and Evolution topic.

7.3.13 Qualitative data were not analysed in the same detail as the quantitative data

Due to time restrictions during the write up of this thesis insufficient time was allocated for the detailed analysis of the qualitative data. The data were coded by hand, whereas the use of a specialist software program such as NVivo would have allowed better data management, reduced the chance of human error/bias when dealing with multiple files and allowed a stronger link between the coded data and the original transcripts. Using a technique such as Iterative coding (IC) would have allowed a more rigorous treatment of the qualitative data from which more valid and potentially repeatable conclusions could have been drawn (Neale, 2016). Rather than being incompatible, the qualitative data collected by this study allowed the exploration of different opinions and additional themes which were not considered in the quantitative data, adding a new complementary dimension to the overall findings of this study.

7.4 Wider Implications of the study

The results of this study have wider implications for the way we train and support non-specialist teachers to enable them to teach evolution confidently in a conceptually accurate manner as well as in the ways we can evaluate the effectiveness of teaching interventions.

7.4.1 Support of non-specialist science teachers and increase in confidence

The results have demonstrated that a relatively short period (60-90 minutes) of focused training can enable non-science specialists to teach natural selection and evolution effectively leading to significantly improved understanding in students of all abilities. The teacher training component of this study may have been a key factor in the recruitment of teachers who wanted to gain a better understanding of the subject matter themselves to facilitate greater understanding in their students. From the feedback and teacher questionnaires we have also shown that participation in the project gave them the support and confidence needed to teach this difficult topic using a variety of different activities and learning styles.

7.4.2 Resources need to be cheap and widely available

This study has shown that simple, cheap and widely available resources can be integrated into carefully planned Schemes of Work. The activities developed for the project can be used in any standard classroom to teach this science topic in a practical and interesting way. Lack of resourcing in science lessons should not be a barrier to learning, rather all children should have an equal opportunity to enjoy and engage in active investigative science lessons.

7.4.3 The replication crisis – can large-scale randomised trail tests in education help?

Although the use of randomised trail tests are mainly confined to the assessment of the relative effectiveness of medical treatments the results of this study builds upon the findings of (Mead et al., 2017) providing further evidence of this method's utility in the quantification and evaluation of different teaching interventions. By combining large scale quantitative data with the qualitative feedback from participants of any study, education researchers can access incredibly rich and diverse sources of information from which they can draw their conclusion with more conviction.

At this juncture it would be prudent to consider 'the replication crisis', an ongoing methodological problem in science, particularly affecting the fields of psychology and medical trials, in which the results of scientific studies cannot be reproduced on subsequent investigation. Reproducibility is an essential part of the scientific method as it; protects against false positives and increases confidence that the result is actually true, the inability to replicate the studies having potentially grave consequences for many fields of science in which significant theories are grounded on unreproducible experimental work. Failure to repeat findings can have several causes including; small sample sizes, changes in circumstances or attitudes over time, sloppy original research/unskilled repeating scientists or in the worse-case scenario – faking. The enormity of the current crisis has been highlighted by Nosek et al. (2015) who showed that ~ 1/3 of psychological studies appearing in premier journals have been able to replicate their findings.

Importantly, this study has presented evidence of reproducibility relating to the utility of the student assessment instrument to assess 9 to 11 year-old students and that the Schemes of Work provided did improve their understanding of evolution. Additionally, reproducibility at the student and school level has been shown with statistically significant correlations with large effect sizes between the rho and P values taken from the individual analyses. However, very few of the results relating to student performance at the teacher/class level were statistically significant in both tranches despite the massive sample sizes, suggesting the underlying explanations for variation in class gain are what seem unrepeatable.

7.5 Closing comments

This study provides the first repeated randomised trial test results from two large-scale tranches of data into the effectiveness of different teaching intervention programmes in 9 to 11 year-old students in the UK. It compares the relative effectiveness of four Schemes of Work developed to help non-science specialists teach the newly introduced Key Stage 2 Inheritance and Evolution topic. Schemes of Work 2 and 3 were found to be more effective relative to the other two programmes of instruction due to the unforeseen reciprocal interaction between pairs of main activities in lessons 2 and 4, exposed by the large sample sizes employed. The teacher centred Peppered moth power point activity is more effective when taught in combination with studying homology in Trilobites whilst, conversely the more student centred ‘hunting’ moths activity is most effective when combined with studying homology in the pentadactyl limb.

This study has provided evidence to support that the student assessment instrument and its mode of delivery are fit for purpose for use with 9 to 11 year-old students. It is of appropriate difficulty allowing clear discrimination between students, its novel mode of delivery mitigates poor literacy skills and enables students to access and understand the difference between alternative responses. The results have demonstrated that the instrument is reliable in multiple ways: there is consistency across both tranches with strong test-retest correlations between the percentage of correct answers per item in the pre-test ($r = 0.98$, Pearson’s correlation coefficient) with no significant difference between the LOESS residual change in score between the pre and pots-tests (P = 0.26, Wilcoxon rank sum test) and across items, with acceptable and improving internal consistency (Cronbach's alpha) scores as the students experienced the test more frequently. The assessment instrument has been shown that it is valid using supporting content related evidence obtained from the feedback and endorsement of the project by participating teachers, with criterion related evidence from the correlation between student scores and their independently assessed ability levels and finally construct related evidence from the analysis of the data itself i.e. after teaching the understanding of the students had improved showing the assessment instrument was testing what it was intending to test.

There is repeatable evidence to suggest that all of the four Schemes of Work are effective at improving understanding as there was significant improvement in student performance in both the post-teaching and retention tests when matched to their pre-test scores, however the magnitude of this improvement varied. The results of this study are directly relevant as the teaching interventions were delivered to students by their normal teachers in their own classrooms and are translatable to being applied by all teachers across the country. Unlike past studies, this present one analyses the effectiveness of an integrated sequence of lessons rather than a single isolated activity.

Finally, far from being a negative result the lack of any significant confounding factors at the teacher and school level after multi test correction is very encouraging. It suggests that all teachers (in all schools) were able to teach natural selection and evolution effectively, regardless of their own understanding and experience level using the resources supplied together with the training and support they received by this project.

In conclusion, this thesis ends with some final comments from participating teachers which sum up the intentions of this project; to help them teach this conceptually difficult topic in a practical and interesting way:

> *"I definitely wouldn't have been able to teach this unit to that level without the resources and that help."*

> *"They really enjoyed the scheme and it wasn't too writing heavy. It included other elements of the curriculum that maybe we don't focus enough on, like cross curriculum links with maths and literacy. That was really good."*

“I found the resources really helpful, and the structure of the scheme, the four lessons followed in a logical order, and for the children as well, they found that sensible. I found the resources really helpful for me because I didn't have a lot of prior knowledge about it, and I found that it was quite easy to use them to teach from. It didn't take loads of time to try and figure out what it all meant before you taught it, it was fairly easy to sort out”.

“Thank you so much, it's been really nice having your involvement and for them to meet you at the end of the project I think was really fabulous. I know you've got so many schools, you probably can't do that all the time, but I think for them they were really interested to meet you at the end and hear about the study.”

“The scheme was good because there were differentiated resources in there.”

“Really, if you hadn't provided the resources, I don't think we could have done the lessons. Obviously just because teachers have so much on their plates, and it's just the ideas, knowing what to do and the preparation. So, thank you for all of that, thank you very much.”

References

Abrie, A., L (2010). Student teachers' attitudes towards and willingness to teach evolution in a changing South African environment. *Journal of Biological Education, 44*(3), 102-107.

Aguillard, D. W. (1998). *An analysis of factors influencing the teaching of biological evolution in Louisiana public secondary schools* (Vol. 60-03). ProQuest Dissertations And Theses; Thesis (Ph.D.)--Louisiana State University and Agricultural & Mechanical College, 1998.: Dissertation Abstracts International.

Ainsworth, S., & Saffer, J. (2013). Can children read evolutionary trees? *Merrill-Palmer Quarterly, 59*(2), 221-247.

Akerson, V., & Donnelly, L. A. (2010). Teaching Nature of Science to K-2 Students: What understandings can they attain? *International Journal of Science Education, 32*(1), 97-124.

Alexander, D. (2014). *Creation or evolution: Do we have to choose?* : Monarch Books.

Allgaier, J. (2009). Scientific Experts and the Controversy About Teaching Creation/Evolution in the UK Press. *Science & Education, 19*(6-8), 797-819. doi:10.1007/s11191-009-9195-5

Allmon, W. D. (2011). Why don't people think evolution is true? Implications for teaching, in and out of the classroom. *Evolution: Education and Outreach, 4*(4), 648-665.

Alters, B. J. (1997). Whose nature of science? *Journal of research in science teaching, 34*(1), 39-55.

Alters, B. J., & Nelson, C. E. (2002). Perspective: Teaching evolution in higher education. *Evolution, 56*(10), 1891-1901.

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of research in science teaching, 39*(10), 952-978.

Anderson, D. L., Fisher, K. M., & Smith, M. U. (2010). Support for the CINS as a diagnostic conceptual inventory: Response to Nehm and Schonfeld (2008). *Journal of research in science teaching, 47*(3), 354-357.

Anderson, L. M., Evertson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first-grade reading groups. *The Elementary School Journal, 79*(4), 193-223.

Andersson, B., & Bach, F. (2005). On designing and evaluating teaching sequences taking geometrical optics as an example. *Science Education, 89*(2), 196-218.

Andrew, M. D., & Schwab, R. L. (1995). Has reform in teacher education influenced teacher performance? An outcome assessment of graduates of an eleven-university consortium. *Action in teacher education, 17*(3), 43-53.

Andrews, J. W. (1980). Preservice Performance and the National Teacher Exams. *Phi Delta Kappan, 61*(5), 358-359.

Archer, J. (1992). Gender stereotyping of school subjects. *The Psychologist, 5*(2), 66-69.

Armbruster, P., Patel, M., Johnson, E., & Weiss, M. (2009). Active learning and student-centered pedagogy improve student attitudes and performance in introductory biology. *CBE-Life Sciences Education, 8*(3), 203-213.

Asghar, A., Wiles, J. R., & Alters, B. (2007). Canadian pre-service elementary teachers' conceptions of biological evolution and evolution education. *McGill Journal of Education, 42*(2), 189-209.

Ashton, P., & Crocker, L. (1987). Systematic study of planned variations: The essential focus of teacher education reform. *Journal of Teacher Education, 38*(3), 2-8.

Asoko, H. (1996). Developing scientific concepts in the primary classroom: Teaching about electric circuits. *Research in science education in Europe*, 36-49.

Ayers, J. B., & Qualls, G. S. (1979). Concurrent and predictive validity of the National Teacher Examinations. *The Journal of Educational Research, 73*(2), 86-92.

Bandoli, J. H. (2008). Do state science standards matter? Comparing student perceptions of the coverage of evolution in Indiana & Ohio public high schools. *The American Biology Teacher, 70*(4), 212-216.

Banet, E., & Ayuso, G. E. (2003). Teaching of biological inheritance and evolution of living beings in secondary school. *International Journal of Science Education, 25*(3), 373-407.

Bao, L. (2006). Theoretical comparisons of average normalized gain calculations. *American Journal of Physics, 74*(10), 917-922.

Barnett, R. C., & Rivers, C. (2004). Men Are From Earth, and So Are Women. It's Faulty Research That Sets Them Apart. *Chronicle of Higher Education, 51*(2), B11-B13.

Barr, R., Dreeben, R., & Wiratchai, N. (1983). *How schools work*: University of Chicago Press.

Barrows, H. S., & Tamblyn, R. M. (1980). *Problem-based learning: An approach to medical education*: Springer Publishing Company.

Bartlett, J. C. (1977). Effects of immediate testing on delayed retrieval: Search and recovery operations with four types of cue. *Journal of Experimental Psychology: Human Learning and Memory, 3*(6), 719.

Basili, P. A., & Sanford, J. P. (1991). Conceptual change strategies and cooperative group work in chemistry. *Journal of research in science teaching, 28*(4), 293-304.

Beard, M. N. C. E. N. J. (1996). Better Biology Teaching by Emphasizing Evolution &the Nature of Science. *The American Biology Teacher, 58*(6), 332-336.

Beardsley, P. M., Bloom, M. V., & Wise, S. B. (2012). Challenges and Opportunities for Teaching and Designing Effective K—12 Evolution Curricula. *Evolution challenges: Integrating research and practice in teaching and learning about evolution*, 287.

Becker, B. J. (1989). Gender and science achievement: A reanalysis of studies from two meta-analyses. *Journal of research in science teaching, 26*(2), 141-169.

Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology, 85*(2), 347.

Berkman, M., & Plutzer, E. (2010). *Evolution, creationism, and the battle to control America's classrooms*: Cambridge University Press.

Berkman, M., & Plutzer, E. (2011). Defeating creationism in the courtroom, but not in the classroom. *Science, 331*(6016), 404-405.

Berkman, M. B., Pacheco, J. S., & Plutzer, E. (2008). Evolution and creationism in America's classrooms: a national portrait. *PLoS biology, 6*(5), e124.

Berkman, M. B., & Plutzer, E. (2015). Enablers of doubt: How future teachers learn to negotiate the evolution wars in their classrooms. *The ANNALS of the American Academy of Political and Social Science, 658*(1), 253-270.

Berlinger, D. (1987). Ways of thinking about students and classrooms by more and less experience teachers. *Exploring teacher's thinking. London, United Kingdom: Cassell.*

Bernard, R., & Clarizio, H. (1981). Socioeconomic bias in special education placement decisions. *Psychology in the Schools, 18*(2), 178-183.

Berti, A. E., Toneatti, L., & Rosati, V. (2010). Children's conceptions about the origin of species: A study of Italian children's conceptions with and without instruction. *The Journal of the Learning Sciences, 19*(4), 506-538.

Betts, J. R., & Shkolnik, J. L. (1999). The behavioral effects of variations in class size: The case of math teachers. *Educational evaluation and policy analysis, 21*(2), 193-213.

Bishop, B. A., & Anderson, C. W. (1986). Evolution by Natural Selection: A Teaching Module. Occasional Paper No. 91.

Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of research in science teaching, 27*(5), 415-427.

Bizzo, N. M. V. (1994). From down house landlord to Brazilian high school students: what has happened to evolutionary knowledge on the way? *Journal of research in science teaching, 31*(5), 537-556.

Blank, L. M., & Andersen, H. O. (1997). Teaching Evolution: Coming To a Classroom Near You? *Reports of the National Center for Science Education, 17*(3), 10-13.
Bloom, P., & Weisberg, D. S. (2007). Childhood origins of adult resistance to science. *Science, 316*(5827), 996-997.
Boggs, C., Watt, W., & Ehrlich, P. (2003). Butterflies: evolution and ecology taking flight: Chicago, IL: University of Chicago Press.
Bonwell, C. C., & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports*: ERIC.
BouJaoude, S., Salloum, S., & Abd-El-Khalick, F. (2004). Relationships between selective cognitive variables and students' ability to solve chemistry problems. *International Journal of Science Education, 26*(1), 63-84.
Bowman, K. L. (2008). The evolution battles in high-school science classes: who is teaching what? *Frontiers in Ecology and the Environment, 6*(2), 69-74. doi:10.1890/070013
Bradley, S., & Taylor, J. (1998). The effect of school size on exam performance in secondary schools. *Oxford bulletin of economics and statistics, 60*(3), 291-324.
Bramley, T., Rodeiro, C., & Vitello, S. (2015). Gender differences in GCSE. *Cambridge Assessment Research Report*.
Bramschreiber, T. L. (2014). Teaching evolution: Strategies for conservative school communities. *Race Equality Teaching, 32*(1), 10-14.
Breakwell, G. M., & Beardsell, S. (1992). Gender, parental and peer influences upon science attitudes and activities. *Public Understanding of Science*.
Brem, S. K., Ranney, M., & Schindel, J. (2003). Perceived consequences of evolution: College students perceive negative personal and social impact in evolutionary theory. *Science Education, 87*(2), 181-206.
Browning, E., & Hohenstein, J. (2013). The use of narrative to promote primary school children's understanding of evolution. *Education 3-13*(ahead-of-print), 1-18.
Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and instruction, 21*(1), 95-108.
Brumby, M. N. (1984). Misconceptions about the concept of natural selection by medical biology students. *Science Education, 68*(4), 493-503.
Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13*(4), 273.
Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & cognition, 36*(3), 604-616.
Butler, A. C., & Roediger III, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*(4-5), 514-527.
Byrne, C. (1983). *Teacher knowledge and teacher effectiveness: A literature review, theoretical analysis and discussion of research strategy.* Paper presented at the meeting of the Northwestern Educational Research association, Ellenville, NY.
Calderhead, J. (1996). Teachers: Beliefs and knowledge.
Campos, R., & Sá-Pinto, A. (2013). Early evolution of evolutionary thinking: teaching biological evolution in elementary schools. *Evolution: Education and Outreach, 6*(1), 25.
Carey, S. (1985). Conceptual change in childhood, 1985. *Bradford, Cambridge, MA*.
Carey, S., & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational psychologist, 28*(3), 235-251.
Carlsen, W. S. (1991). Effects of new biology teachers' subject-matter knowledge on curricular planning. *Science Education, 75*(6), 631-647.
Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & cognition, 20*(6), 633-642.

Carter, B. E., & Wiles, J. R. (2014). Scientific consensus and social controversy: exploring relationships between students' conceptions of the nature of science, biological evolution, and global climate change. *Evolution: Education and Outreach, 7*(1), 6.

Carter, K., Sabers, D., Cushing, K., Pinnegar, S., & Berliner, D. C. (1987). Processing and using information about students: A study of expert, novice, and postulant teachers. *Teaching and Teacher Education, 3*(2), 147-157.

Catley, K. M., & Novick, L. R. (2008). Seeing the wood for the trees: an analysis of evolutionary diagrams in biology textbooks. *BioScience, 58*(10), 976-987.

Cavallo, A. M. (1996). Meaningful learning, reasoning ability, and students' understanding and problem solving of topics in genetics. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 33*(6), 625-656.

Cavallo, A. M., & McCall, D. (2008). Seeing may not mean believing: examining students' understandings & beliefs in evolution. *The American Biology Teacher, 70*(9), 522-530.

Centola, D., Wilensky, U., & McKenzie, E. (2000). Survival of the groupiest: Facilitating students' understanding of the multiple levels of fitness through multi-agent modeling-The EACH Project. *International Journal of Complex Systems, 377*.

Chanet, B., & Lusignan, F. (2008). Teaching Evolution in Primary Schools: An Example in French Classrooms. *Evolution: Education and Outreach, 2*(1), 136-140. doi:10.1007/s12052-008-0095-y

Cheek, K. A. (2013). Exploring the Relationship between Students' Understanding of Conventional Time and Deep (Geologic) Time. *International Journal of Science Education, 35*(11), 1925-1945.

Chi, M. T. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *The Journal of the Learning Sciences, 14*(2), 161-199.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73-105.

Chinn, C. A., & Samarapungavan, A. (2001). Distinguishing between understanding and belief. *Theory into Practice, 40*(4), 235-241.

Chinsamy, A., & Plagányi, É. (2008). Accepting evolution. *Evolution, 62*(1), 248-254.

Chuang, H. C. (2003). Teaching evolution: attitudes & strategies of educators in Utah. *The American Biology Teacher, 65*(9), 669-674.

Cid, C. R. (2013). The effects of a professional development program on elementary and middle school teachers' understanding and acceptance of macroevolution and how they teach it.

Cobern, W. W. (1994). Comments and criticism. Point: Belief, understanding, and the teaching of evolution. *Journal of research in science teaching, 31*(5), 583-590.

Colburn, A. I. (1994). Misconceptions in evolution education. *Investigating Evolutionary Biology in the Laboratory*, 11-14.

Coley, J. D., & Tanner, K. D. (2012). Common origins of diverse misconceptions: cognitive principles and the development of biology thinking. *CBE-Life Sciences Education, 11*(3), 209-215.

Collins, F. S. (2006). *The language of God: A scientist presents evidence for belief*: Simon and Schuster.

Connolly, P., Keenan, C., & Urbanska, K. (2018). The trials of evidence-based practice in education: a systematic review of randomised controlled trials in education research 1980–2016. *Educational Research, 60*(3), 276-291.

Cook, L., & Saccheri, I. (2012). The peppered moth and industrial melanism: evolution of a natural selection case study. *Heredity, 110*(3), 207-212.

Cooper, J. L., Robinson, P., & Ball, D. A. (2003). The interactive lecture: Reconciling group and active learning strategies with traditional instructional formats. *Exchanges, the Online Journal of Teaching and Learning in the CSU*.

Cooper, R. A. (2001). The goal of evolution instruction: Belief or literacy. *Reports of the National Center for Science Education, 21*(1-2), 14-18.

Crawford, B. A. (1999). Is it realistic to expect a preservice teacher to create an inquiry-based classroom? *Journal of Science Teacher Education, 10*(3), 175-194.
Crawford, B. A., Zembal-Saul, C., Munford, D., & Friedrichsen, P. (2005). Confronting prospective teachers' ideas of evolution and scientific inquiry using technology and inquiry-based tasks. *Journal of research in science teaching, 42*(6), 613-637.
Croll, P., & Moses, D. (2000). Ideologies and utopias: Education professionals' views of inclusion. *European Journal of Special Needs Education, 15*(1), 1-12.
Cunningham, D. J., & Anderson, R. C. (1968). Effects of practice time within prompting and confirmation presentation procedures on paired associate learning. *Journal of Verbal Learning and Verbal Behavior, 7*(3), 613-616.
Cunningham, D. L., & Wescott, D. J. (2009). Still more “fancy” and “myth” than “fact” in students’ conceptions of evolution. *Evolution: Education and Outreach, 2*(3), 505-517.
Curry, A. (2009). Creationist beliefs persist in Europe: American Association for the Advancement of Science.
Dagher, Z. R., & Boujaoude, S. (2005). Students' perceptions of the nature of evolutionary theory. *Science Education, 89*(3), 378-391.
Darwin, C. (2004). *On the origin of species, 1859*: Routledge.
Davson-Galle, P. (2004). Understanding:‘knowledge’,‘belief’and ‘understanding’. *Science & Education, 13*(6), 591-598.
Dawkins, R. (1986). The blind watchmaker
. *UK: Penguin Books*.
Dawkins, R. (2006). The God Delusion New York. *Google Scholar*.
Deemer, S. (2004). Classroom goal orientation in high school classrooms: Revealing links between teacher beliefs and classroom environments. *Educational Research, 46*(1), 73-90.
Deller, S. C., & Rudnicki, E. (1993). Production efficiency in elementary education: The case of Maine public schools. *Economics of education review, 12*(1), 45-57.
Demastes, S. S., Settlage, J., & Good, R. (1995). Students' conceptions of natural selection and its role in evolution: Cases of replication and comparison. *Journal of research in science teaching, 32*(5), 535-550.
Dempster, E. R., & Hugo, W. (2006). Introducing the concept of evolution into South African schools. *South African Journal of Science, 102*.
Deniz, H., Çetin, F., & Yılmaz, I. (2011). Examining the relationships among acceptance of evolution, religiosity, and teaching preference for evolution in Turkish preservice biology teachers. *Reports of the National Center for Science Education, 31*(4).
Deniz, H., & Donnelly, L. A. (2011). Preservice secondary science teachers’ acceptance of evolutionary theory and factors related to acceptance. *Reports of the National Center for Science Education, 31*(4).
Deniz, H., Donnelly, L. A., & Yilmaz, I. (2008). Exploring the factors related to acceptance of evolutionary theory among Turkish preservice biology teachers: Toward a more informative conceptual ecology for biological evolution. *Journal of research in science teaching, 45*(4), 420-443.
Dennett, D. C. (1995). Darwin's dangerous idea. *The Sciences, 35*(3), 34-40.
Denton, J. J., & Peters, W. H. (1988). Program Assessment Report Curriculum Evaluation of a Non-traditional Program for Certifying Teachers.
Denzin, N. K. (2012). Triangulation 2.0. *Journal of mixed methods research, 6*(2), 80-88.
Dobzhansky, T. (1973). <Dobzhansky T 1973 Nothing in Biology makes sense except in the liight of Evolution.pdf>. *Am Biol Teach, 35*, 125-129.
Dodick, J., & Orion, N. (2003). Cognitive factors affecting student understanding of geologic time. *Journal of research in science teaching, 40*(4), 415-442.
Dole, J. A., & Sinatra, G. M. (1998). Reconceptalizing change in the cognitive construction of knowledge. *Educational psychologist, 33*(2-3), 109-128.

Dotger, S., Dotger, B. H., & Tillotson, J. (2010). Examining how preservice science teachers navigate simulated parent–teacher conversations on evolution and intelligent design. *Science Education, 94*(3), 552-570.
Dougherty, M. (2011). Six million years ago, what set our ancestors on the path from ape to human? *The American Biology Teacher, 73*(2), 66-66.
Doyle, W., & Carter, K. (2003). Narrative and learning to teach: Implications for teacher-education curriculum. *Journal of Curriculum Studies, 35*(2), 129-137.
Driver, R. (1989). Students' conceptions and the learning of science. *International Journal of Science Education, 11*(5), 481-490.
Driver, R. (1994). *Making sense of secondary science: Support materials for teachers*: Psychology Press.
Driver, R., Asoko, H., Leach, J., Scott, P., & Mortimer, E. (1994). Constructing scientific knowledge in the classroom. *Educational researcher, 23*(7), 5-12.
Driver, R., Leach, J., & Millar, R. (1996). *Young people's images of science*: McGraw-Hill Education (UK).
Driver, R., Leach, J., Scott, P., & Wood-Robinson, C. (1994). Young people's understanding of science concepts: Implications of cross-age studies for curriculum planning.
Duit, R., Treagust, D., & Widodo, A. (2008). Teaching science for conceptual change: Theory and practice *International handbook of research on conceptual change* (pp. 629-646): Routledge.
Ebert-May, D., Brewer, C., & Allred, S. (1997). Innovation in large lectures: Teaching for active learning. *BioScience, 47*(9), 601-607.
Eccles, J. (1985). *Sex differences in achievement patterns.* Paper presented at the Nebraska symposium on motivation.
Eccles, J. S., & Blumenfeld, P. (1985). Classroom experiences and student gender: Are there differences and do they matter? *Gender influences in classroom interaction* (pp. 79-114): Elsevier.
Eccles, J. S., Wigfield, A., Midgley, C., Reuman, D., Iver, D. M., & Feldlaufer, H. (1993). Negative effects of traditional middle schools on students' motivation. *The Elementary School Journal, 93*(5), 553-574.
Elhoweris, H., Mutua, K., Alsheikh, N., & Holloway, P. (2005). Effect of children's ethnicity on teachers' referral and recommendation decisions in gifted and talented programs. *Remedial and Special Education, 26*(1), 25-31.
Elley, W. B. (1966). The role of errors in learning with feedback. *British Journal of Educational Psychology, 36*(3), 296-300.
Elwood, J., & Comber, C. (1995). *Gender differences in 'A'level examinations: the reinforcement of stereotypes.* Paper presented at the Paper presented as part of the symposium-A New ERA.
Emmons, N. A., & Kelemen, D. A. (2015). Young children's acceptance of within-species variation: Implications for essentialism and teaching evolution. *Journal of experimental child psychology, 139*, 148-160.
Evans, E. M. (2000). The emergence of beliefs about the origins of species in school-age children. *Merrill-Palmer Quarterly (1982-)*, 221-254.
Evans, E. M. (2001). Cognitive and contextual factors in the emergence of diverse belief systems: Creation versus evolution. *Cognitive psychology, 42*(3), 217-266.
Evans, E. M. (2008). Conceptual change and evolutionary biology: A developmental analysis. *International handbook of research on conceptual change*, 263-294.
Evans, E. M., Spiegel, A. N., Gram, W., Frazier, B. N., Tare, M., Thompson, S., & Diamond, J. (2010). A conceptual guide to natural history museum visitors' understanding of evolution. *Journal of research in science teaching, 47*(3), 326-353.
Evans, E. M., Weiss, M., Lane, J. D., & Palmquist, S. (2015). The Spiral Model. *Cognitive Development in Museum Settings: Relating Research and Practice*, 36.
Fail, J. (2008). A no-holds-barred evolution curriculum for elementary and junior high school students. *Evolution: Education and Outreach, 1*(1), 56-64.

Farber, N. (2006). Conducting qualitative research: A practical guide for school counselors. *Professional school counseling, 9*(4), 367-375.
Feiman-Nemser, S. (2001). From Preparation to Practice: Designing a Continuum To Strengthen and Sustain Teaching. *Teachers College Record, 103*(6), 1013-1055.
Ferrari, M., & Chi, M. T. (1998). The nature of naive explanations of natural selection. *International Journal of Science Education, 20*(10), 1231-1256.
Fetherstonhaugh, T., & Treagust, D. F. (1992). Students' understanding of light and its properties: Teaching to engender conceptual change. *Science Education, 76*(6), 653-672.
Fine, G. A., & Sandstrom, K. L. (1988). *Knowing children: Participant observation with minors* (Vol. 15): Sage Newbury Park, CA.
Finn, J. D., & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27*(3), 557-577.
Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational evaluation and policy analysis, 21*(2), 97-109.
Fisher, A. V., Godwin, K. E., & Seltman, H. (2014). Visual Environment, Attention Allocation, and Learning in Young Children When Too Much of a Good Thing May Be Bad. *Psychol Sci, 25*(7), 1362-1370.
Fisher, K. M. (2004). The importance of prior knowledge in college science instruction. *Reform in undergraduate science teaching for the 21st century*, 69-83.
Flanagan, J. C., & Roseman, J. E. (2011). *Assessing middle and high school students' understanding of evolution with standards-based items.* Paper presented at the National Association of Research in Science Teaching Annual Meeting, Orlando, FL.
Fowler Jr, W. J., & Walberg, H. J. (1991). School size, characteristics, and outcomes. *Educational evaluation and policy analysis, 13*(2), 189-202.
Fowler, S. R., & Meisels, G. G. (2010). Florida teachers' attitudes about teaching evolution. *The American Biology Teacher, 72*(2), 96-99.
Fraenkel, J., & Wallen, R. (2010). N, E. (2010) How to design and Evaluate Research in Education: New York: McGraw-Hill.
Frasier, T. R., & Roderick, C. (2011). Improving how evolution is taught: facilitating a shift from memorization to evolutionary thinking. *Evolution: Education and Outreach, 4*(2), 298-307.
Friedrichsen, P., Driel, J. H. V., & Abell, S. K. (2011). Taking a closer look at science teaching orientations. *Science Education, 95*(2), 358-376.
Galli, L. M. G., & Meinardi, E. N. (2011). The role of teleological thinking in learning the Darwinian model of evolution. *Evolution: Education and Outreach, 4*(1), 145-152.
Garvin-Doxas, K., & Klymkowsky, M. W. (2008). Understanding randomness and its impact on student learning: lessons learned from building the Biology Concept Inventory (BCI). *CBE-Life Sciences Education, 7*(2), 227-233.
Gelman, S. A., & Rhodes, M. (2012). 'Two-thousand years of stasis. *Evolution challenges: Integrating research and practice in teaching and learning about evolution*, 3.
Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155-170.
Geraedts, C. L., & Boersma, K. T. (2006). Reinventing natural selection. *International Journal of Science Education, 28*(8), 843-870.
Gerrig, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading*: Yale University Press.
Gilbert, J. K., Osborne, R. J., & Fensham, P. J. (1982). Children's science and its consequences for teaching. *Science Education, 66*(4), 623-633.
Glass, G. V. (1982). School class size: Research and policy.
Glaze, A. L., & Goldston, M. J. (2015). US Science Teaching and Learning of Evolution: A Critical Review of the Literature 2000–2014. *Science Education, 99*(3), 500-518.
Glaze, A. M. L. (2013). *Evolution and pre-service science teachers: investigating acceptance and rejection*: The University of Alabama.

Glover, J. A. (1989). The" testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81*(3), 392.
Goldston, M. J., & Kyzer, P. (2009). Teaching evolution: Narratives with a view from three southern biology teachers in the USA. *Journal of research in science teaching, 46*(7), 762-790.
Gould, S., & Haq, S. N. (1999). Rocks of Ages: Science and Religion in the Fullness of Life. *Nature, 400*(6747), 830-830.
Gould, S. J. (1981). Evolution as fact and theory. *Discover, 2*(5), 34-37.
Gould, S. J. (1999). Dorothy, it's really Oz :A pro-creationist decision in Kansas is more than a blow against Darwin,. *Time Magazine*.
Grace, M., Hanley, P., & Johnson, S. (2008). ‘Darwin-inspired’science: teachers’ views, approaches and needs. *School Science Review, 90*(331), 71-77.
Greany, T., Barnes, I., Mostafa, T., Pensiero, N., & Swensson, C. (2016). Trends in Maths and Science Study (TIMSS): National Report for England.
Green, L. (1993). Possible gender bias within teachers' perceptions of pupils with special needs. *Support for Learning, 8*(2), 77-80.
Greene, E. D. (1990). The logic of university students' misunderstanding of natural selection. *Journal of research in science teaching, 27*(9), 875-885.
Greene, J. C. (1987). Stakeholder participation in evaluation design: Is it worth the effort? *Evaluation and Program planning, 10*(4), 379-394.
Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational evaluation and policy analysis, 11*(3), 255-274.
Gregory, T. R. (2009). Understanding natural selection: essential concepts and common misconceptions. *Evolution: Education and Outreach, 2*(2), 156-175.
Griffith, J. A., & Brem, S. K. (2004). Teaching evolutionary biology: Pressures, stress, and coping. *Journal of research in science teaching, 41*(8), 791-809.
Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*: Teachers College Press, Teachers College, Columbia University.
Gunn, A. M. (2004). *Evolution and Creationism in the Public Schools: A Handbook for Educators, Parents and Community Leaders*: McFarland.
Ha, M., Haury, D. L., & Nehm, R. H. (2012). Feeling of certainty: Uncovering a missing link between knowledge and acceptance of evolution. *Journal of research in science teaching, 49*(1), 95-121.
Hagay, S. a. (2005). High school students’ conceptions related to speciation and common descent]. *Unpublished raw data*.
Hall, G. E., & Hord, S. M. (2006). Implementing change: Patterns, principles, and potholes.
Halldén, O. (1988). The evolution of the species: pupil perspectives and school perspectives. *International Journal of Science Education, 10*(5), 541-552.
Haney, W., Madaus, G., & Kreitzer, A. (1987). Chapter 5: Charms Talismanic: Testing Teachers for the Improvement of American Education. *Review of research in education, 14*(1), 169-238.
Harlen, W. (1997). Primary teachers' understanding in science and its impact in the classroom. *Research in Science Education, 27*(3), 323-337.
Harlen, W., & Holroyd, C. (1997). Primary teachers’ understanding of concepts of science: Impact on confidence and teaching. *International Journal of Science Education, 19*(1), 93-105.
Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development, 77*(3), 505-524.
Harvey, T., & Edwards, P. (1980). Children's expectations and realisations of science. *British Journal of Educational Psychology, 50*(1), 74-76.
Hashweh, M. Z. (2005). Teacher pedagogical constructions: a reconfiguration of pedagogical content knowledge. *Teachers and Teaching, 11*(3), 273-292.
Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*: Routledge.

Heck, R. H., & Mayor, R. A. (1993). School characteristics, school academic indicators and student outcomes: Implications for policies to improve schools. *Journal of Education Policy, 8*(2), 143-154.
Heller, K. A., & Ziegler, A. (1996). Gender differences in mathematics and the sciences: Can attributional retraining improve the performance of gifted females? *Gifted child quarterly, 40*(4), 200-210.
Hermann, R. S. (2008). Evolution as a controversial issue: a review of instructional approaches. *Science & Education, 17*(8-9), 1011-1032.
Hermann, R. S. (2011). Breaking the cycle of continued evolution education controversy: On the need to strengthen elementary level teaching of evolution. *Evolution: Education and Outreach, 4*(2), 267-274.
Hermann, R. S. (2013). High school biology teachers’ views on teaching evolution: implications for science teacher educators. *Journal of Science Teacher Education, 24*(4), 597-616.
Herreid, C. F. (1998). Why isn't cooperative learning used to teach science? *BioScience, 48*(7), 553-559.
Hewson, M. G., & Hewson, P. W. (1983). Effect of instruction using students' prior knowledge and conceptual change strategies on science learning. *Journal of research in science teaching, 20*(8), 731-743.
Hillis, D. M. (2007). Making evolution relevant and exciting to biology students. *Evolution, 61*(6), 1261-1264.
Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10*(5), 562-567.
Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*(3), 297-313.
Hokayem, H., & BouJaoude, S. (2008). College students' perceptions of the theory of evolution. *Journal of research in science teaching, 45*(4), 395-419.
Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581.
Hyde, J. S., & Linn, M. C. (2006). Gender similarities in mathematics and science. *Science, 314*(5799), 599-600.
Järnefelt, E., Canfield, C. F., & Kelemen, D. (2015). The divided mind of a disbeliever: Intuitive beliefs about nature as purposefully created among different groups of non-religious adults. *Cognition, 140*, 72-88.
Jarvis, T., Pell, A., & McKeon, F. (2003). Changes in primary teachers' science knowledge and understanding during a two year in-service programme. *Research in Science & Technological Education, 21*(1), 17-42.
Jeffery, W. R. (2008). Emerging model systems in evo-devo: cavefish and microevolution of development. *Evolution & development, 10*(3), 265-272.
Jensen, M. S., & Finley, F. N. (1995). Teaching evolution using historical arguments in a conceptual change strategy. *Science Education, 79*(2), 147-166.
Jensen, M. S., & Finley, F. N. (1996). Changes in students' understanding of evolution resulting from different curricular and instructional strategies. *Journal of research in science teaching, 33*(8), 879-900.
Jesky-Smith, R. (2002). Me, teach science? *Science and Children, 39*(6), 26.
Jiménez-Aleixandre, M. P. (1992). Thinking about theories or thinking with theories?: a classroom study with natural selection. *International Journal of Science Education, 14*(1), 51-61.
Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college what evidence is there that it works? *Change: the magazine of higher learning, 30*(4), 26-35.
Jovanovic, J., & King, S. S. (1998). Boys and Girls in the Performance-Based Science classroom: Who’s doing the performing? *American Educational Research Journal, 35*(3), 477-496.

Kaess, W., & Zeaman, D. (1960). Positive and negative knowledge of results on a Pressey-type punchboard. *Journal of Experimental Psychology, 60*(1), 12.
Kagan, D. M. (1990). Ways of evaluating teacher cognition: Inferences concerning the Goldilocks principle. *Review of Educational Research, 60*(3), 419-469.
Kahan, D. M. (2015). Climate-science communication and the measurement problem. *Political Psychology, 36*(S1), 1-43.
Kahle, J. B., & Lakes, M. K. (1983). The myth of equality in science classrooms. *Journal of research in science teaching, 20*(2), 131-140.
Kampourakis, K., & Zogza, V. (2008). Students' intuitive explanations of the causes of homologies and adaptations. *Science & Education, 17*(1), 27-47.
Kampourakis, K., & Zogza, V. (2009). Preliminary evolutionary explanations: A basic framework for conceptual change and explanatory coherence in evolution. *Science & Education, 18*(10), 1313-1340.
Kaplan, R., & Saccuzzo, D. (1997). Psychological testing: Principles, Applications and Issues (4e éd.). Pacific Grove: Brooks: Cole Publishing (1re éd. 1982).
Kargbo, D. B., HOBBS, E. D., & ERICKSON, G. L. (1980). Children's beliefs about inherited characteristics. *Journal of Biological Education, 14*(2), 137-146.
Kelemen, D. (1999a). The scope of teleological thinking in preschool children. *Cognition, 70*(3), 241-272.
Kelemen, D. (1999b). Why are rocks pointy? Children's preference for teleological explanations of th natural world. *Developmental psychology, 35*(6), 1440.
Kelemen, D. (2012). Teleological minds: How natural intuitions about agency and purpose influence learning about evolution. *Evolution challenges: Integrating research and practice in teaching and learning about evolution, 66*, 92.
Kelemen, D., Emmons, N. A., Seston Schillaci, R., & Ganea, P. A. (2014). Young children can be taught basic natural selection using a picture-storybook intervention. *Psychol Sci, 25*(4), 893-902. doi:10.1177/0956797613516009
Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition, 111*(1), 138-143.
Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*(1), 1-24.
Keown, D. (1988). Teaching evolution: Improved approaches for unprepared students. *The American Biology Teacher*, 407-410.
Kikas, E. (2004). Teachers' conceptions and misconceptions concerning three natural phenomena. *Journal of research in science teaching, 41*(5), 432-448.
Kim, S. Y., & Nehm, R. H. (2011). A Cross-Cultural Comparison of Korean and American Science Teachers' Views of Evolution and the Nature of Science. *International Journal of Science Education, 33*(2), 197-227.
Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. Retrieved from
Klitgaard, R. E., & Hall, G. R. (1975). Are there unusually effective schools? *Journal of Human Resources*, 90-106.
Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychol Sci, 15*(10), 694-698.
Krueger, A. B. (1999). Experimental estimates of education production functions. *The quarterly journal of economics, 114*(2), 497-532.
Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *The Economic Journal, 111*(468), 1-28.
Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education, 77*(3), 319-337.
Lamdin, D. J. (1995). Testing for the effect of school size on student achievement within a school district. *Education Economics, 3*(1), 33-42.

Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2009). Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Medical education, 43*(12), 1174-1181.

Lawson, A. E. (1983). Predicting science achievement: The role of developmental level, disembedding ability, mental capacity, prior knowledge, and beliefs. *Journal of research in science teaching, 20*(2), 117-129.

Lawson, A. E., & Thompson, L. D. (1988). Formal reasoning ability and misconceptions concerning genetics and natural selection. *Journal of research in science teaching, 25*(9), 733-746.

Lawson, A. E., & Weser, J. (1990). The rejection of nonscientific beliefs about life: Effects of instruction and reasoning skills. *Journal of research in science teaching, 27*(6), 589-606.

Lawson, A. E., & Worsnop, W. A. (1992). Learning about evolution and rejecting a belief in special creation: Effects of reflective reasoning skill, prior knowledge, prior belief and religious commitment. *Journal of research in science teaching, 29*(2), 143-166.

Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of research in science teaching, 29*(4), 331-359.

Legare, C. H., Lane, J. D., & Evans, E. M. (2013). Anthropomorphizing science: how does it affect the development of evolutionary concepts? *Merrill-Palmer Quarterly, 59*(2), 168-197.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41*(3), 635-679.

Levesque, P. J., & Guillaume, A. M. (2010). Teachers, evolution, and religion: No resolution in sight. *Review of Religious Research*, 349-365.

Livingston, C., & Borko, H. (1989). Expert-novice differences in teaching: A cognitive analysis and implications for teacher education. *Journal of Teacher Education, 40*(4), 36-42.

Lombrozo, T., Thanukos, A., & Weisberg, M. (2008). The importance of understanding the nature of science for accepting evolution. *Evolution: Education and Outreach, 1*(3), 290-298.

Lord, T., & Marino, S. (1993). How university students view the theory of evolution. *Journal of College Science Teaching, 22*(6), 353-357.

Losh, S. C., & Nzekwe, B. (2011). The influence of education major: How diverse preservice teachers view pseudoscience topics. *Journal of Science Education and Technology, 20*(5), 579-591.

Lovely, E. C., & Kondrick, L. C. (2008). Teaching evolution: challenging religious preconceptions. *Integr Comp Biol, 48*(2), 164-174. doi:10.1093/icb/icn026

Machin, S., & McNally, S. (2005). Gender and student achievement in English schools. *Oxford review of economic policy, 21*(3), 357-372.

Majerus, M. E. (2009). Industrial melanism in the peppered moth, Biston betularia: an excellent teaching example of Darwinian evolution in action. *Evolution: Education and Outreach, 2*(1), 63-74.

Mancebon, M. J., & Molinero, C. M. (2000). Performance in primary schools. *Journal of the Operational Research Society, 51*(7), 843-854.

Marek, E. A. (2008). Why the learning cycle? *Journal of Elementary Science Education, 20*(3), 63.

Marek, E. A., & Cavallo, A. M. (1997). *The learning cycle: Elementary school science and beyond*: ERIC.

Martin, J. W. (2010). Compatibility of major US Christian denominations with evolution. *Evolution: Education and Outreach, 3*(3), 420-431.

Martin, M. O., Mullis, I. V., Foy, P., & Stanco, G. M. (2012). *TIMSS 2011 International Results in Science*: ERIC.

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Blunk, M., Crawford, B., Kelly, B., & Meyer, K. M. (1994). Enacting project-based science: Experiences of four middle grade teachers. *The Elementary School Journal, 94*(5), 517-538.

Marzano, R. J. (2003). *What works in schools: Translating research into action*: ASCD.
Mayer, R. E. (2002). Multimedia learning *Psychology of learning and motivation* (Vol. 41, pp. 85-139): Elsevier.
Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*: Harvard University Press.
Mayr, E. (1991a). The ideological resistance to Darwin's theory of natural selection. *Proceedings of the American Philosophical Society, 135*(2), 123-139.
Mayr, E. (1991b). *One long argument: Charles Darwin and the genesis of modern evolutionary thought* (Vol. 2): Harvard University Press.
Mayr, E. (1992). The idea of teleology. *Journal of the History of Ideas*, 117-135.
Mayr, E. (1997). *Evolution and the diversity of life: selected essays*: Harvard University Press.
Mayr, E. (2001). The philosophical foundations of Darwinism. *Proceedings of the American Philosophical Society, 145*(4), 488-495.
Mazur, A. (2004). Believers and disbelievers in evolution. *Politics and the Life Sciences, 23*(2), 55-61.
McComas, W. F. (1998). The principal elements of the nature of science: Dispelling the myths *The nature of science in science education* (pp. 53-70): Springer.
McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*(4-5), 494-513.
McKeachie, W. J., Lin, Y.-G., & Strayer, J. (2002). Creationist vs. evolutionary beliefs: Effects on learning biology. *The American Biology Teacher, 64*(3), 189-192.
McVaugh, N. K., Birchfield, J., Lucero, M. M., & Petrosino, A. J. (2011). Evolution education: Seeing the forest for the trees and focusing our efforts on the teaching of evolution. *Evolution: Education and Outreach, 4*(2), 286-292.
Mead, L. S., & Scott, E. C. (2010a). Problem concepts in evolution part I: purpose and design. *Evolution: Education and Outreach, 3*(1), 78-81.
Mead, L. S., & Scott, E. C. (2010b). Problem concepts in evolution part II: cause and chance. *Evolution: Education and Outreach, 3*(2), 261-264.
Mead, R., Hejmadi, M., & Hurst, L. D. (2017). Teaching genetics prior to teaching evolution improves evolution understanding but not acceptance. *PLoS biology, 15*(5), e2002255.
Meadows, L., Doster, E., & Jackson, D. F. (2000). Managing the conflict between evolution & religion. *The American Biology Teacher, 62*(2), 102-107.
Meikle, W. E., & Scott, E. C. (2010). Why are there still monkeys? *Evolution: Education and Outreach, 3*(4), 573-575.
Metz, K. E. (2010). *Scaffolding children's understanding of the fit between organisms and their environment in the context of the practices of science.* Paper presented at the Proceedings of the 9th International Conference of the Learning Sciences-Volume 1.
Miller, J. D., Scott, E. C., & Okamoto, S. (2006). Public acceptance of evolution. *SCIENCE-NEW YORK THEN WASHINGTON-, 313*(5788), 765.
Miller, K. R. (2008). *Only a theory: Evolution and the battle for America's soul*: Penguin.
Millis, B. J. (2002). Enhancing Learning--and More!--Through Cooperative Learning. IDEA Paper.
Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational evaluation and policy analysis, 21*(2), 165-177.
Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of education review, 13*(2), 125-145.
Moore, R. (2002). Teaching evolution: do state standards matter? *AIBS Bulletin, 52*(4), 378-381.
Moore, R., & Cotner, S. (2009). Rejecting Darwin: the occurrence & impact of creationism in high school biology classrooms. *The American Biology Teacher, 71*(2), e1-e4.
Morrison, J. A., & Lederman, N. G. (2003). Science teachers' diagnosis and understanding of students' preconceptions. *Science Education, 87*(6), 849-867.

Mpeta, M. (2014). *The influence of the beliefs of teachers and learners on the teaching and learning of evolution.*

Murnane, R. J., & Phillips, B. R. (1981). Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teaching performance. *Economics of education review, 1*(4), 453-465.

Murphy, C., Neil, P., & Beggs, J. (2007). Primary science teacher confidence revisited: ten years on. *Educational Research, 49*(4), 415-430.

Nadelson, L. (2009). Preservice teacher understanding and vision of how to teach biological evolution. *Evolution: Education and Outreach, 2*(3), 490-504.

Nadelson, L., Culp, R., Bunn, S., Burkhart, R., Shetlar, R., Nixon, K., & Waldron, J. (2009). Teaching Evolution Concepts to Early Elementary School Students. *Evolution: Education and Outreach, 2*(3), 458-473. doi:10.1007/s12052-009-0148-x

Nadelson, L., & Hardy, K. (2015). Trust in science and scientists and the acceptance of evolution. *Evolution: Education and Outreach, 8 (9)*, 1-9.

Nadelson, L., & Nadelson, S. (2010). K-8 educators perceptions and preparedness for teaching evolution topics. *Journal of Science Teacher Education, 21*(7), 843-858.

Nadelson, L., & Sinatra, G. M. (2008). Educational professionals' knowledge and acceptance of evolution. *Evolutionary Psychology*.

Nadelson, L. S., & Sinatra, G. M. (2010). Shifting acceptance of evolution: promising evidence of the influence of the Understanding Evolution website. *The Researcher, 23*(1), 13-29.

Nadelson, L. S., & Southerland, S. A. (2010). Examining the interaction of acceptance and understanding: How does the relationship change with a focus on macroevolution? *Evolution: Education and Outreach, 3*(1), 82-88.

Neale, D. C., Smith, D., & Johnson, V. G. (1990). Implementing conceptual change teaching in primary science. *The Elementary School Journal, 91*(2), 109-131.

Neale, J. (2016). Iterative categorization (IC): a systematic technique for analysing qualitative data. *Addiction, 111*(6), 1096-1106.

Nehm, R. H., Kim, S. Y., & Sheppard, K. (2009). Academic preparation in biology and advocacy for teaching evolution: Biology versus non-biology teachers. *Science Education, 93*(6), 1122-1146.

Nehm, R. H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. *BioScience, 57*(3), 263-272.

Nehm, R. H., & Schonfeld, I. S. (2007). Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching of evolution in schools? *Journal of Science Teacher Education, 18*(5), 699-723.

Nehm, R. H., & Schonfeld, I. S. (2008). Measuring knowledge of natural selection: A comparison of the CINS, an open-response instrument, and an oral interview. *Journal of research in science teaching, 45*(10), 1131-1160.

Nehm, R. H., & Schonfeld, I. S. (2010). The future of natural selection knowledge measurement: a reply to Anderson et al.(2010). *Journal of research in science teaching, 47*(3), 358-362.

Nelson, C. E. (2008). Teaching evolution (and all of biology) more effectively: strategies for engagement, critical reasoning, and confronting misconceptions. *Integr Comp Biol, 48*(2), 213-225.

Nelson, C. E. (2012). Why don't undergraduates really "get" evolution? What can faculty do. *Evolution challenges: Integrating research and practice in teaching and learning about evolution. New York: Oxford University Press. p*, 311-347.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., . . . Christensen, G. (2015). Promoting an open research culture. *Science, 348*(6242), 1422-1425.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., . . . Gonsalkorale, K. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences, 106*(26), 10593-10597.

Novak, J. D., Mintzes, J. I., & Wandersee, J. H. (2005). Learning, teaching, and assessment: A human constructivist perspective *Assessing science understanding* (pp. 1-13): Elsevier.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational evaluation and policy analysis, 21*(2), 127-142.

Ohlsson, S. (1991). Young adults' understanding of evolutionary explanations: preliminary observations. *Thecnical Report. Learning Research and Development Center, University of Pittsburgh*.

Oliva, J. M. (2003). The structural coherence of students' conceptions in mechanics and conceptual change. *International Journal of Science Education, 25*(5), 539-561.

Osborne, J., & Simon, S. (1996). Primary science: past and future directions.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049-1079.

Padian, K. (2013). Correcting some common misrepresentations of evolution in textbooks and the media. *Evolution: Education and Outreach, 6*(1), 11.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307-332.

Palmer, D. H. (1999). Exploring the link between students' scientific and nonscientific conceptions. *Science Education, 83*(6), 639-653.

Papadopoulou, P., Stanissavljevic, J., Katakos, E., & Athanasiou, K. (2011). *Acceptance and understanding of evolution theory: A comparative study of Greek and Serbian teachers.* Paper presented at the Science learning and Citizenship, E-book Proceedings of the ESERA 2011 Conference, Lyon, France. http://lsg. ucy. ac. cy/esera/e_book/base/index. html.

Pardhan, H., & Bano, Y. (2001). Science teachers' alternate conceptions about direct-currents. *International Journal of Science Education, 23*(3), 301-318.

Passmore, C., & Stewart, J. (2002). A modeling approach to teaching evolutionary biology in high schools*. *Journal of research in science teaching, 39*(3), 185-204.

Penner, D. E. (2000). Explaining systems: Investigating middle school students' understanding of emergent phenomena. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 37*(8), 784-806.

Petrosino, A. J., Lucero, M. M., & Mann, M. J. (2015). Decentralized thinking and understanding of evolution in K-12 evolution education. *Evolution: Education and Outreach, 8*(1), 2.

Piaget, J. (1964). Part I: Cognitive development in children: Piaget development and learning. *Journal of research in science teaching, 2*(3), 176-186.

Piaget, J., & Cook, M. (1952). *The origins of intelligence in children* (Vol. 8): International Universities Press New York.

Pigott, R. L., & Cowen, E. L. (2000). Teacher race, child race, racial congruence, and teacher ratings of children's school adjustment. *Journal of School Psychology, 38*(2), 177-195.

Pobiner, B. (2016). Accepting, understanding, teaching, and learning (human) evolution: Obstacles and opportunities. *American journal of physical anthropology, 159*(S61), S232-S274.

Poling, D. A., & Evans, E. M. (2002). Why do birds of a feather flock together? Developmental change in the use of multiple explanations: Intention, teleology and essentialism. *British Journal of Developmental Psychology, 20*(1), 89-112.

Poling, D. A., & Evans, E. M. (2004a). Are dinosaurs the rule or the exception?: Developing concepts of death and extinction. *Cognitive development, 19*(3), 363-383.

Poling, D. A., & Evans, E. M. (2004b). Religious belief, scientific expertise, and folk ecology. *Journal of Cognition and Culture, 4*(3), 485-524.

Poole, M. (2008). Creationism, intelligent design and science education. *School Science Review, 90*(330), 123-129.
Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research, 2004*(121), 63-73.
Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211-227.
Prince, M. (2004). Does active learning work? A review of the research. *Journal of engineering education, 93*(3), 223-231.
Prinou, L., Halkia, L., & Skordoulis, C. (2008). What conceptions do Greek school students form about biological evolution? *Evolution: Education and Outreach, 1*(3), 312-317.
Quirk, T. J., Witten, B. J., & Weinberg, S. F. (1973). Review of Studies of the Concurrent and Predictive Validity of the National Teacher Examinations 1. *Review of Educational Research, 43*(1), 89-113.
Rector, M. A., Nehm, R. H., & Pearl, D. (2013). Learning the language of evolution: Lexical ambiguity and word meaning in student explanations. *Research in Science Education, 43*(3), 1107-1133.
Reiss, M. (1992). How should science teachers teach the relationship between science and religion. *School Science Review, 74*(267), 126-130.
Reiss, M. J. (2008). Should science educators deal with the science/religion issue? *Studies in Science Education, 44*(2), 157-186.
Remmers, H., & Remmers, E. M. (1926). The negative suggestion effect on true-false examination questions. *Journal of Educational Psychology, 17*(1), 52.
Resnick, M. (1996). Beyond the centralized mindset. *The Journal of the Learning Sciences, 5*(1), 1-22.
Rice, J. K. (1999). The impact of class size on instructional strategies and the use of time in high school mathematics and science courses. *Educational evaluation and policy analysis, 21*(2), 215-229.
Rice, J. W., Clough, M. P., Olson, J. K., Adams, D. C., & Colbert, J. T. (2015). University faculty and their knowledge & acceptance of biological evolution. *Evolution: Education and Outreach, 8*(1), 8.
Richards, R. A. (2008). Philosophical challenges in teaching evolution. *Evolution: Education and Outreach, 1*(2), 158-164.
Rissler, L. J., Duncan, S. I., & Caruso, N. M. (2014). The relative importance of religion and education on university students' views of evolution in the Deep South and state science standards across the United States. *Evolution: Education and Outreach, 7*(1), 24.
Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458.
Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychol Sci, 17*(3), 249-255.
Rosenholtz, S. (1986). The organizational context of teaching: In learning to teach. *Urbana-Champaign, Illinois: University of Illinois at Urbana-Champaign*.
Rosenholtz, S. J., & Kyle, S. J. (1984). Teacher isolation: Barrier to professionalism. *American Educator, 8*(4), 10-15.
Rossman, G. B., & Wilson, B. L. (1985). Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study. *Evaluation review, 9*(5), 627-643.
Royal Society. (2006). Royal Society statement on evolution, creationism and intelligent design. http://www.royalsoc.ac.uk/news.asp?id=4298, *(accessed June 2017).*
Rudolph, J. L., & Stewart, J. (1998). Evolution and the nature of science: On the historical discord and its implications for education. *Journal of research in science teaching, 35*(10), 1069-1089.
Rutledge, M. L. (1996). Indiana high school biology teachers and evolutionary theory: Acceptance and understanding.

Rutledge, M. L., & Mitchell, M. A. (2002). High school biology teachers' knowledge structure, acceptance & teaching of evolution. *The American Biology Teacher, 64*(1), 21-28.
Rutledge, M. L., & Sadler, K. C. (2007). Reliability of the measure of acceptance of the theory of evolution (MATE) instrument with university students. *The American Biology Teacher, 69*(6), 332-335.
Rutledge, M. L., & Sadler, K. C. (2011). University Students' Acceptance of Biological Theories-Is Evolution Really Different? *Journal of College Science Teaching, 41*(2), 38.
Rutledge, M. L., & Warden, M. A. (1999). The development and validation of the measure of acceptance of the theory of evolution instrument. *School Science and Mathematics, 99*(1), 13-18.
Rutledge, M. L., & Warden, M. A. (2000). Evolutionary theory, the nature of science & high school biology teachers: Critical relationships. *The American Biology Teacher, 62*(1), 23-31.
Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*(4), 1054-1070.
Samarapungavan, A., & Wiers, R. W. (1997). Children's thoughts on the origin of species: A study of explanatory coherence. *Cognitive Science, 21*(2), 147-177.
Sander, W. (1993). Expenditures and student achievement in Illinois: New evidence. *Journal of Public Economics, 52*(3), 403-416.
Sanders, L. R., Borko, H., & Lockard, J. D. (1993). Secondary science teachers' knowledge base when teaching science courses in and out of their area of certification. *Journal of research in science teaching, 30*(7), 723-736.
Sanders, M. (2010). *Meeting the curriculum requirement of" learner-centredness" when teaching evolution: Giving learners and a fair deal.* Paper presented at the Proceedings, 17 th annual conference of the Southern African Researchers in Mathematics, Science and Technology Education.
Sanders, M., & Ngxola, N. (2009). Addressing teachers' concerns about teaching evolution. *Journal of Biological Education, 43*(3), 121-128. doi:10.1080/00219266.2009.9656166
Sanders, W. L., & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of personnel evaluation in education, 8*(3), 299-311.
Saunders, W. L. (1992). The constructivist perspective: Implications and teaching strategies for science. *School Science and Mathematics, 92*(3), 136-141.
Schank, R. C. (1995). *What we learn when we learn by doing*. Retrieved from
Scharmann, L. C. (2005). A proactive strategy for teaching evolution. *The American Biology Teacher, 67*(1), 12-16.
Scott, E. C. (2000). Not (just) in Kansas anymore. *Science, 288*(5467), 813-815.
Settlage, J. (1994). Conceptions of natural selection: A snapshot of the sense-making process. *Journal of research in science teaching, 31*(5), 449-457.
Shankar, G., & Skoog, G. D. (1993). Emphasis given evolution and creationism by Texas high school biology teachers. *Science Education, 77*(2), 221-233.
Sharp, L. M., & Frankel, J. (1983). Respondent burden: A test of some common assumptions. *Public Opinion Quarterly, 47*(1), 36-53.
Shortland, R., & Mark, M. (1987). Improving inferences from multiple methods. *Multiple Methods in Program Evaluation. New directions for program evaluation*(35).
Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. *Cognitive psychology, 52*(2), 170-194.
Shtulman, A., & Calabi, P. (2012). Cognitive constraints on the understanding and acceptance of evolution. *Evolution challenges: Integrating research and practice in teaching and learning about evolution*, 47-65.
Shtulman, A., & Calabi, P. (2013). Tuition vs. intuition: Effects of instruction on naive theories of evolution. *Merrill-Palmer Quarterly, 59*(2), 141-167.

Shtulman, A., Neal, C., & Lindquist, G. (2016). Children's ability to learn evolutionary explanations for biological adaptation. *Early Education and Development, 27*(8), 1222-1236.
Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science, 32*(6), 1049-1062.
Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review, 57*(1), 1-23.
Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher, 15*(2), 4-14.
Shulman, L. S., & Grossman, P. (1988). Knowledge growth in teaching: A final report to the Spencer Foundation. *Palo Alto, Calif.: Stanford University*.
Sickel, A. J., & Friedrichsen, P. (2013). Examining the evolution education literature with a focus on teachers: major findings, goals for teacher preparation, and directions for future research. *Evolution: Education and Outreach, 6*(1), 23.
Sinatra, G. M., Brem, S. K., & Evans, E. M. (2008). Changing Minds? Implications of Conceptual Change for Teaching and Learning about Biological Evolution. *Evolution: Education and Outreach, 1*(2), 189-195. doi:10.1007/s12052-008-0037-8
Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demastes, J. W. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of research in science teaching, 40*(5), 510-528.
Sinclair, B. B., Naizer, G., & Ledbetter, C. (2011). Observed implementation of a science professional development program for K-8 classrooms. *Journal of Science Teacher Education, 22*(7), 579-594.
Smith, E. L., Blakeslee, T. D., & Anderson, C. W. (1993). Teaching strategies associated with conceptual change learning in science. *Journal of research in science teaching, 30*(2), 111-126.
Smith, M. U. (1994). Counterpoint: Belief, Understanding, and the Teaching of Evolution. *Journal of research in science research, 31*(5), 591-597.
Smith, M. U. (2010a). Current status of research in teaching and learning evolution: I. Philosophical/epistemological issues. *Science & Education, 19*(6-8), 523-538.
Smith, M. U. (2010b). Current status of research in teaching and learning evolution: II. Pedagogical issues. *Science & Education, 19*(6-8), 539-571.
Smith, M. U., & Scharmann, L. C. (1999). Defining versus describing the nature of science: A pragmatic analysis for classroom teachers and science educators. *Science Education, 83*(4), 493-509.
Smith, M. U., & Siegel, H. (2004). Knowing, believing, and understanding: What goals for science education? *Science & Education, 13*(6), 553-582.
Solomon, G. E. (2002). Birth, kind and naïve biology. *Developmental Science, 5*(2), 213-218.
Solomon, G. E., & Johnson, S. C. (2000). Conceptual change in the classroom: Teaching young children to understand biological inheritance. *British Journal of Developmental Psychology, 18*(1), 81-96.
Solomon, J. (1994). The rise and fall of constructivism.
Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of research in science teaching, 46*(6), 610-631.
Southerland, S. A., Abrams, E., Cummins, C. L., & Anzelmo, J. (2001). Understanding students' explanations of biological phenomena: Conceptual frameworks or p-prims? *Science Education, 85*(4), 328-348.
Southerland, S. A., Sinatra, G. M., & Matthews, M. R. (2001). Belief, knowledge, and science education. *Educational Psychology Review, 13*(4), 325-351.
Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 30*(9), 641.
Sprietsma, M. (2013). Discrimination in grading: experimental evidence from primary school teachers. *Empirical economics, 45*(1), 523-538.

Springer, K. (1992). Children's awareness of the biological implications of kinship. *Child Development, 63*(4), 950-959.
Springer, K. (1995). Acquiring a naive theory of kinship through inference. *Child Development, 66*(2), 547-558.
Springer, K. (1996). Young Children's Understanding of a Biological Basis for Parent-Offspring Relations. *Child Development, 67*(6), 2841-2856.
Springer, L., Stanne, M. E., & Donovan, S. (1999). Measuring the success of small-group learning in college level SMET teaching: a meta-analysis. *Review of Educational Research, 69*(9), 21-51.
Stasz, C., & Stecher, B. M. (2000). Teaching mathematics and language arts in reduced size and non-reduced size classrooms. *Educational evaluation and policy analysis, 22*(4), 313-329.
Stebbins, R. C., & Allen, B. (1975). Simulating evolution. *The American Biology Teacher, 37*(4), 206-211.
Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743.
Summers, A. A., & Wolfe, B. L. (1975). Which School Resources Help Learning? Efficiency and Equity in Philadelphia Public Schools. Business Review.
Summers, M., & Kruger, C. (1994). A longitudinal study of a constructivist approach to improving primary school teachers' subject matter knowledge in science. *Teaching and Teacher Education, 10*(5), 499-519.
Tamir, P., & Zohar, A. (1991). Anthropomorphism and teleology in reasoning about biological phenomena. *Science Education, 75*(1), 57-67.
Tanner, K., & Allen, D. (2005). Approaches to biology teaching and learning: understanding the wrong answers—teaching toward conceptual change. *Cell Biology Education, 4*(2), 112-117.
Tashakkori, A., & Creswell, J. W. (2007). The new era of mixed methods: Sage Publications.
Taylor, M., Cartwright, B. S., & Bowden, T. (1991). Perspective taking and theory of mind: Do children predict interpretive diversity as a function of differences in observers' knowledge? *Child Development, 62*(6), 1334-1351.
Taylor, N., & Coll, R. (1997). The use of analogy in the teaching of solubility to pre-service primary teachers. *Australian Science Teachers Journal, 43*(4), 58.
Teddlie, C., & Tashakkori, A. (2003). Major issues and controveries inthe use of mixed methods in the social and behvioral sciences. *Handbook of mixed methods in social & behavioral research, Thousand Oaks, CA: Sage.*, 3-50.
Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*: Sage.
Tekkaya, C., Akyol, G., & Sungur, S. (2012). Relationships among Teachers' Knowledge and Beliefs Regarding the Teaching of Evolution: A Case for Turkey. *Evolution: Education and Outreach, 5*(3), 477-493.
Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*(2), 253.
Thagard, P., & Findlay, S. (2009). Getting to Darwin: Obstacles to Accepting Evolution by Natural Selection. *Science & Education, 19*(6-8), 625-636. doi:10.1007/s11191-009-9204-8
Thijs, G. D. (1992). Evaluation of an introductory course on ‚force' considering students' preconceptions. *Science Education, 76*(2), 155-174.
Tidon, R., & Lewontin, R. C. (2004). Teaching evolutionary biology. *Genetics and molecular biology, 27*(1), 124-131.
Tobin, K. G., Kahle, J. B., & Fraser, B. J. (1990). *Windows into science classrooms: Problems associated with higher-level cognitive learning*: Psychology Press.
Trani, R. (2004). I won't teach evolution; it's against my religion. And now for the rest of the story…. *The American Biology Teacher, 66*(6), 419-427.

Trumper, R., & Gorsky, P. (1993). Learning about energy: The influence of alternative frameworks, cognitive levels, and closed-mindedness. *Journal of research in science teaching, 30*(7), 637-648.
Udovic, D., Morris, D., Dickman, A., Postlethwait, J., & Wetherwax, P. (2002). Workshop biology: demonstrating the effectiveness of active learning in an introductory biology course. *AIBS Bulletin, 52*(3), 272-281.
Urquiza-Haas, E. G., & Kotrschal, K. (2015). The mind behind anthropomorphic thinking: attribution of mental states to other species. *Animal behaviour, 109*, 167-176.
Van de Walle, J. A. (1990). *Elementary School Mathematics, Teaching Developmentally*: ERIC.
Van der Riet, M., & Durrheim, K. (2006). Putting design into practice: Writing and evaluating research proposals. *Research in practice: Applied methods for the social sciences, 2*, 80-111.
van Dijk, E. M., & Kattmann, U. (2009). Teaching evolution with historical narratives. *Evolution: Education and Outreach, 2*(3), 479-489.
Vance, V. S., & Schlechty, P. C. (1982). The distribution of academic ability in the teaching force: Policy implications. *The Phi Delta Kappan, 64*(1), 22-27.
Veal, W. R., & Kubasko Jr, D. S. (2003). Biology and Geology Teachers' Domain-Specific Pedagogical Content Knowledge of Evolution. *Journal of Curriculum and Supervision, 18*(4), 334-352.
Venville, G., & Donovan, J. (2005). Searching for clarity to teach the complexity of the gene concept.
Venville, G., & Donovan, J. (2007). Developing year 2 students' theory of biology with concepts of the gene and DNA. *International Journal of Science Education, 29*(9), 1111-1131.
Venville, G., Gribble, S. J., & Donovan, J. (2005). An exploration of young children's understandings of genetics concepts from ontological and epistemological perspectives. *Science Education, 89*(4), 614-633.
Wagler, R. (2010). A missing link: K-4 biological evolution content standards. *Evolution: Education and Outreach, 3*(3), 443-450.
Wagler, R. (2012). Assessing “the framework” for kindergarten through fifth grade biological evolution. *Evolution: Education and Outreach, 5*(2), 274-278.
Walker, J., Cotner, S. H., Baepler, P. M., & Decker, M. D. (2008). A delicate balance: integrating active learning into a large lecture course. *CBE-Life Sciences Education, 7*(4), 361-367.
Weinburgh, M. (1995). Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991. *Journal of research in science teaching, 32*(4), 387-398.
Weiss, M., & Dreesmann, D. C. (2014). Aspirations and Expectations: Comparing Scientist and Teacher Views as a Source of Ideas for Teaching Evolution. *Universal Journal of Educational Research, 2*(5), 421-431.
Werth, A. (2012). Avoiding the pitfall of progress and associated perils of evolutionary education. *Evolution: Education and Outreach, 5*(2), 249-265.
Werth, A. J. (2009). Clearing the Highest Hurdle: Human-Based Case Studies Broaden Students' Knowledge of Core Evolutionary Concepts. *Journal of Effective Teaching, 9*(2), 38-53.
Werth, A. J. (2013). An Evolutionary Focus Improves Students’ Understanding of All Biology. *Reports of the National Center for Science Education, 33*(1).
Wescott, D. J., & Cunningham, D. L. (2005). Recognizing student misconceptions about science and evolution. *Age, 22*(92), 23-29.
Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory, 11*(6), 571-580.
Wheeler, M. A., & Roediger III, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychol Sci, 3*(4), 240-246.

Wilensky, U., & Novak, M. (2010). Teaching and learning evolution as an emergent process: The BEAGLE project. *Epistemology and science education: Understanding the evolution vs. intelligent design controversy*, 213-243.
Wilensky, U. J. (1993). *Connected mathematics: builiding concrete relationships with mathematical knowledge.* Massachusetts Institute of Technology.
Wiles, J. R., & Alters, B. (2011). Effects of an educational experience incorporating an inventory of factors potentially influencing student acceptance of biological evolution. *International Journal of Science Education, 33*(18), 2559-2585.
Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594.
Williams, J. D. (2008). Creationist teaching in school science: A UK perspective. *Evolution: Education and Outreach, 1*(1), 87-95.
Williams, J. D. (2009). Belief versus acceptance: Why do people not believe in evolution? *BioEssays, 31*(11), 1255-1262.
Williams, J. D. (2013). “It’s just a theory”: trainee science teachers’ misunderstandings of key scientific terminology. *Evolution: Education and Outreach, 6*(1), 12.
Williams, K. A., & Cavallo, A. M. (1995). Reasoning Ability, Meaningful Learning, and Students' Understanding of Physics Concepts: Relating Students' Reasoning Ability and Learning Styles To Their Physics Misconceptions. *Journal of College Science Teaching*, 311-314.
Williams, M., DeBarger, A. H., Montgomery, B. L., Zhou, X., & Tate, E. (2012). Exploring middle school students' conceptions of the relationship between genetic inheritance and cell division. *Science Education, 96*(1), 78-103.
Wilson, S. M., Shulman, L. S., & Richert, A. E. (1987). " 150 different ways" of knowing: Representations of knowledge in teaching.
Wood-Robinson, C. (1994). Young people's ideas about inheritance and evolution.
Woods, C. S., & Scharmann, L. C. (2001). High school students’ perceptions of evolutionary theory. *Electronic Journal of Science Education, 6*(2), 1-9.
Yasri, P., & Mancy, R. (2014). Understanding student approaches to learning evolution in the context of their perceptions of the relationship between science and religion. *International Journal of Science Education, 36*(1), 24-45.
Yates, T. B., & Marek, E. A. (2013). Is Oklahoma really OK? A regional study of the prevalence of biological evolution-related misconceptions held by introductory biology teachers. *Evolution: Education and Outreach, 6*(1), 6.
Yates, T. B., & Marek, E. A. (2014). Teachers teaching misconceptions: a study of factors contributing to high school biology students’ acquisition of biological evolution-related misconceptions. *Evolution: Education and Outreach, 7*(1), 7.
Yip, D. Y. (1998). Teachers' misconceptions of the circulatory system. *Journal of Biological Education, 32*(3), 207-215.
Zimmerman, M. (2010). The Clergy Letter Project.
Zuckerman, P. (2005). Atheism: Contemporary Rates and Patterns, chapter in The Cambridge Companion to Atheism, ed. by Michael Martin: Cambridge University Press.

Appendices

Appendix A

Existing studies researching the teaching and learning of evolution in primary age children.

Authors of paper	Sample size and context	Research method	Topic studied
(Kargbo et al., 1980)	32 Canadian children	Clinical interviews but no hypothesis tested	Acquisition of inherited traits
(Springer, 1992)	72 American children aged between 4 and 7	Structured oral interviews using cue cards	Ability to distinguish between biological and non-biological relationships
(Springer, 1995)	54 American children aged between 3 and 5	Structured oral interviews	Understanding that babies grow inside the mother and may inherit properties form her
(Springer, 1996)	121 American children aged between 4 and 7	Structured oral interviews	Resemblance of family members and that shared properties don't always entail kinship
(Samarapungavan & Wiers, 1997)	35 Dutch children aged between 8 and 12.	Oral interview based on semi structured questionnaire	Speciation
(Solomon & Johnson, 2000)	64 American children aged between 5-6.	Structured oral interviews	Understanding of biological inheritance
(Evans, 2001)	102 American children aged between 5-13. (Christian fundamentalist and non-fundamentalist.)	Structured interviews and parental questionnaires	Origins of evolutionist/antievolutionist beliefs
(Solomon, 2002)	28 American preschool children aged between 4 and 5	Oral interviews using case studies and visual prompts	Resemblance of family members based on race and apparel
(Poling & Evans, 2004a)	68 American children aged between 4-9	Structured oral interview using photographs	Concepts of death and extinction
(Venville et al., 2005)	90 Australian children aged between 9 and 15	Interviews with qualitative data collection, followed by quantitative scoring	Understanding of basic inheritance and molecular genetics concepts involved in the theory of kinship
(Venville & Donovan, 2007)	17 Australian 6-7 year -old children	Oral interviews and analysis of reflective journals	Development of students' theory of biology with concepts of the gene and DNA based on intervention lessons
(Chanet & Lusignan, 2008)	French children aged 4-11	Analysis of lesson products	Designed 5 activities to introduce evolution and enhance scientific reasoning skills
(Shtulman & Schulz, 2008)	43 American children aged between 4-9	Interviews structured around picture book	Identification of intraspecific variation
(Berti et al., 2010)	39 Italian children aged between 7 and 9	Structured oral interviews	Conceptions about the origin of species
(Browning & Hohenstein, 2013)	62 English children aged between 5 and 8	Semi-structured interview oral interview	Use of narrative to promote understanding of evolution in primary school children

(Legare et al., 2013)	88 American children aged between 5-11	Structured interviews on narrative recall of stories	Influence of language used to explain evolutionary concepts on understanding evolutional change
(Kelemen et al., 2014)	61 American children aged between 5 and 8	Structured oral interviews consisting of open and closed questions	Explanation of natural selection through story book intervention
(Järnefelt et al., 2015)	40 American children aged between 5-8	Interviews based on forced choice questions on picture cue cards about fictional species.	Essentialism and intraspecific variation
(Shtulman et al., 2016)	96 American children aged 4-12	Tutorials and interviews based around a picture book resource	Evolutionary explanations for biological adaptation

Appendix B

Breakdown of student questionnaire question concept together with original Flanagan and Roseman (2011) and adapted readability scores for the GEVO2teach project.

F & R Ref.	Question	Subject	FRE Flan	FKRG Flan	FRE GEVO	FKRG GEVO
EN046004	1	Common ancestor (extant)	65.9	6.3	66.7	5.6
EN047003	2	Common ancestor (extant/extinct)	60.9	7.6	62.6	7.6
EN05002	3	Common ancestor (animal/plants)	67.7	5.8	62.8	6.8
ENO23001	4	Natural selection (lizards)	69.4	7.2	75.4	5.8
EN028002	5	Natural selection (birds beaks)	56.3	8.3	80.3	5.9
EN032002	6	Natural selection (origin of intraspecific variation)	65.3	6.2	66.9	6.2
EN038003	7	Natural selection (requirements)	64.6	6.6	64.6	6.6
EN039002	8	Intra specific variation (chance of survival)	46.9	8.5	66.7	5.6
EN045002	9	Intra specific variation (chance of survival)	59.2	7.6	68.2	6.2
EN013002	10	Fossils (what a fossil is)	74.9	7.8	77.1	4.4
EN014002	11	Extinction (pattern of extinction)	59.4	7	63.2	6.5
EN052001	12	Using fossils	60.4	8.5	80.8	4.1
EN054002	13	Inter specific variation (lizards and oak trees)	59.5	7.7	77.1	4.4
EN055001	14	Geological time (existence of species)	75.4	5.5	75.4	5.5
EN017002	15	Geological time (changes in environmental conditions)	62.3	7	62.8	7
		MEAN	63.21	7.17	70.04	5.88

Appendix C

C1 Original full written version of Student Questionnaire

<u>Student Questionnaire</u>

<u>Instructions :</u>

- Please fill in the front of this booklet.
- This assessment has 2 sections.
- There are 15 multiple-choice questions and then a table about an adoption story to complete.
- Answer all of the questions in the booklet.
- Please do not work with anyone else or share answers
- Don't worry if you don't know an answer.

Your Name: ____________________.

Your Date of Birth (including year)___________________

Are you a Boy or a Girl?____

Name of your School: _____________________________

Name of Science Teacher: _______________________

Date:____________________

- Your answers will be kept strictly confidential during this study.
- Your answers will not affect your science grades at school.
- By completing this questionnaire, you will be helping research by the Evolution Education Trust and the University of Bath.

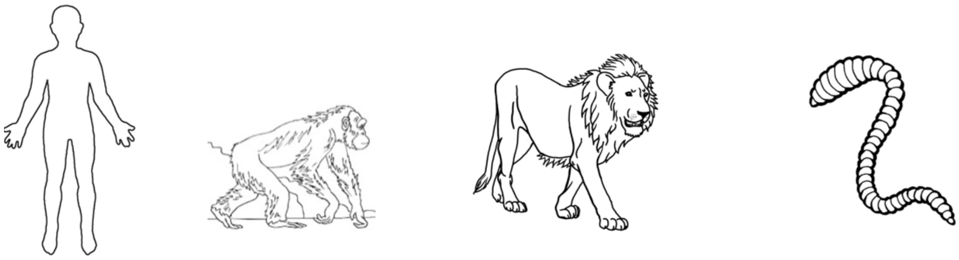
Thank you for your help

THE EVOLUTION EDUCATION TRUST

UNIVERSITY OF BATH

For each question please circle <u>ONE</u> letter to show the correct answer.

1.



Some organisms, like a human and a chimpanzee (chimp) have many similarities. Others like a lion and a worm have fewer. Which answer is TRUE about the ancestors of all these organisms?

A. Humans and chimps share a common ancestor. But lions and worms don't share a common ancestor with each other.
B. Humans and chimps share a common ancestor. Lions and worms share a common ancestor. But humans and chimps do not share a common ancestor with lions and worms
C. None of them share a common ancestor
D. All of them share an ancient common ancestor

2. Which answer is TRUE about species living today and extinct species?

A. A living species and an extinct species that have similarities could share a common ancestor. But if they have few similarities they couldn't share one.
B. A living species and an extinct species could share a common ancestor that lived a long time ago. Even if they have few similarities.
C. No living species shares a common ancestor with an extinct species.
D. Living species could share a common ancestor with each other. But extinct species couldn't share a common ancestor with each other.

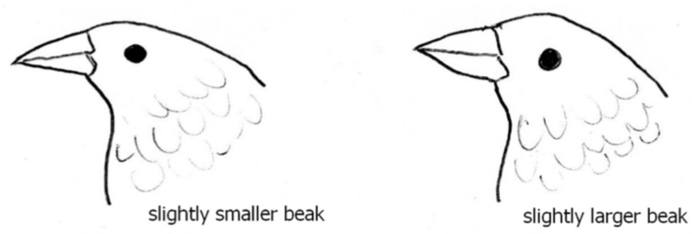
3. Which of these answers is TRUE about the evolution of plants and animals?

A. All plants and animals share a common ancestor with each other.
B. All plants share a common ancestor. But all animals don't share a common ancestor.
C. All animals share a common ancestor. But all plants don't share a common ancestor.
D. None of them share a common ancestor.

4. According to the theory of natural selection, what would happen to a species of lizards when a new predator appears where the lizards live?

A. Lizards that already have traits that help them to avoid the predator would be more likely to survive and reproduce. The ones that don't would be less likely to survive and reproduce.
B. All lizards would try to develop new traits to avoid the new predator.
C. Some lizards would try to develop new traits to avoid the new predator, and the others would die.
D. All the lizards have the same traits because they are from the same species. None of them can survive and reproduce better than anyone else. They would all either die or survive.

5. A species of bird ate many types of seeds. The birds' beaks were different (varied) in size.



The climate changed to much drier conditions and the only seeds that were left were much larger ones. After many generations almost all of the birds had slighter larger beaks.

A. The birds with bigger beaks were better at eating the larger seeds. Only big beaked birds got enough food to survive, reproduce and pass the big beak trait on to the next generation.
B. Birds with smaller beaks had to work really hard to open the bigger seeds. The more they used their beaks the larger they grew. They were then able to get enough food to survive, reproduce and pass on the big beak trait.
C. Birds with smaller beaks grew their beaks so they could open the seeds to get enough food to survive, reproduce and pass on the big beak trait.
D. It was chance that all the birds' beaks got bigger in one generation. They could then get enough food to survive, reproduce and pass on the big beak trait.

6. Could individuals of species look different today than individuals of the same species did many generations ago?

A. YES. All individuals can change a little and pass on those changes to their offspring.
B. YES. Some individuals can change a little and pass on those changes to their offspring.
C. YES. Some individuals born with certain traits are more likely to survive and pass on these traits to their offspring.
D. NO. Species do not change even after many generations. So individuals of the same species wouldn't look different.

7. Which of these is NEEDED for the process of natural selection to occur?

A. Members of the same species must compete with each other.
B. Members of different species must compete with each other.
C. There must be a sudden environmental change.
D. Traits must be inherited from one generation to the next

8. Individual members of a species have differences in inherited traits. What could these differences change? **Select one answer from the table below.**

Answer	Their ability to find food	Their ability to attract a mate
A	✓	✓
B	✓	✗
C	✗	✓
D	✗	✗

9. Which answer is TRUE about individuals of the same species?

A. They may have different inherited traits. These differences may change the individual's chance to survive and reproduce.
B. They would have the same inherited traits because they are from the same species. So they all have an equal chance of surviving and reproducing.
C. Individuals of the same species have the same inherited traits. But they pick up different skills and learn new things during their lifetime. Only these picked up traits cause differences in survival and reproduction.
D. They may have different inherited traits. These different traits do not change the individual's chance of survival and reproduction.

10. Which answer best describes what a fossil can be? **Select one answer form the table below.**

Answer	A bone in which the original matter has been replaced by rock	An impression left by a bone in rock
A	✓	✗
B	✗	✓
C	✓	✓
D	✗	✗

11. Which answer is TRUE about extinction of species?

A. Very few species have ever become extinct. Most are still alive.
B. There have been extinction events when many species became extinct at about the same time. Apart from these events, extinction is very rare.
C. Humans have caused the majority of extinctions. Up until recently species rarely became extinct.
D. Many species have become extinct throughout the history of life on earth.

12. A scientist finds some fossils of an extinct species of fish. What could she do by studying the fossils? **Select one answer from the table below.**

Answer	Discover what features the extinct species had	Discover similarities and differences between features of the extinct species and those alive today
A	✓	✓
B	✓	✗
C	✗	✓
D	✗	✗

13. Which answer is TRUE about oak trees and lizards?

A. There are similarities and differences between oak trees and lizards.
B. There are similarities between oak trees and lizards but no differences.
C. There are differences between oak trees and lizards but no similarities.
D. There is no way to tell if oak trees and lizards have similarities or differences.

14. Which answer is TRUE about the species that are living on earth today?

A. All species living today have existed since the time life began.
B. Most species living today have existed since the time life began. But a few have appeared more recently.
C. Most species alive today didn't exist at the time life began.
D. There is no way of finding out. You can't tell whether all, most or a few species living today have existed since the time life began.

15. Which answer is TRUE about how environmental conditions have changed since life began on earth?

A. Conditions have remained about the same everywhere on earth. Only minor changes from year to year.
B. Conditions have remained the same in the oceans but have changed on land.
C. Conditions have remained the same except for a few sudden changes in certain places. Such as a meteorite hitting the earth.
D. Conditions have changed dramatically. Some of these changes have happened suddenly and others more gradually.

The adoption story

There was a woman called Mrs. Smith. Mrs. Smith went into hospital and gave birth to a baby girl. Unfortunately, Mrs. Smith died just after having the baby and she never even got to see or hold her baby. Fortunately, there was a really nice lady called Mrs. Jones who was visiting the hospital. Mrs. Jones had always wanted a little girl and saw the baby girl was all alone after her birth mother died. Mrs. Jones adopted the baby girl and took her home immediately to live with her. Mrs. Jones loved the little girl very much and called her "Daughter" and the little girl loved her very much and called her "Mummy". The little girl spent her whole life with Mrs. Jones.

Think about who the little baby girl was born to? Who did she live with and grow up with?

Now think about what the little girl is going to be like when she grows up into a young woman. Fill in the column for the grown up adopted baby girl. Will she be like Mrs. Smith or Mrs. Jones for each different feature?

Feature	Mrs. Smith (Birth mother)	Mrs. Jones (Adoptive mother)	Grown up adopted baby girl
Skin colour	White	Black	
Hair colour	Ginger	Brown	
Favourite TV soap	EastEnders	Coronation Street	
Nose shape	Turned up at end	Straight	
Tattoo	Butterfly on shoulder	None	
Shape of pancreas (an organ inside the body)	Flat	Round	
Chin dimple	Doesn't have one	Does have one	
Eating habits	Vegetarian	Not vegetarian	
Going to church	Didn't go to church	Goes to church every week	
Freckles	Has freckles	Doesn't have freckles	

Now explain your answers in as much detail as possible with what you know about variation.

C2 Amended written version of Student Questionnaire

Student Questionnaire

Instructions:

- Please fill in the front of this booklet.
- This assessment has 2 sections.
- There are 15 multiple-choice questions and then a table about an adoption story to complete.
- Answer all of the questions in the booklet.
- Please do not work with anyone else or share answers
- Don't worry if you don't know an answer.

Your Name: ___________________.

Your Date of Birth (including year)__________________

Are you a Boy or a Girl?____

Name of your School: ___________________________

Name of Science Teacher: ______________________

Date:___________________

Pre	Post	Retention

- Your answers will be kept strictly confidential during this study.
- Your answers will not affect your science grades at school.
- By completing this questionnaire you will be helping research by the Evolution Education Trust and the University of Bath.

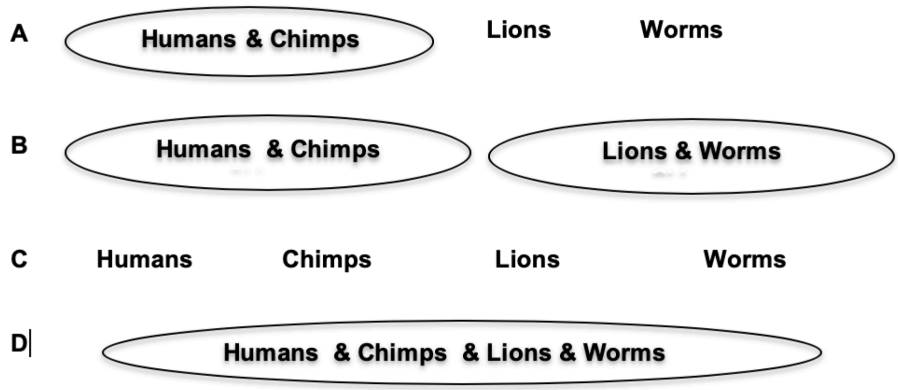
Thank you for your help

THE EVOLUTION EDUCATION TRUST

UNIVERSITY OF BATH

For each question please circle ONE letter to show the correct answer. A trait = a characteristic e.g. eye colour.

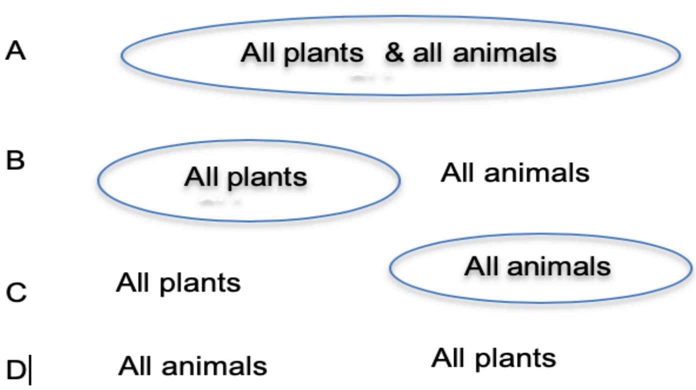
1. Some organisms, like a human and a chimpanzee (chimp) have many similarities. Others like a lion and a worm have fewer. The circles show organisms that share a common ancestor. **Which answer is TRUE? Select one answer from the options below.**



2. Which answer is **TRUE** about a species living today and an extinct species?

   A. A living species and an extinct one can share a common ancestor, but only if they have lots of similarities.
   B. A living species and an extinct one can share a common ancestor that lived a long time ago. Even if they have few similarities.
   C. No living species shares a common ancestor with an extinct one.
   D. Only living species can share a common ancestor with each other. Extinct species can't share a common ancestor with each other.

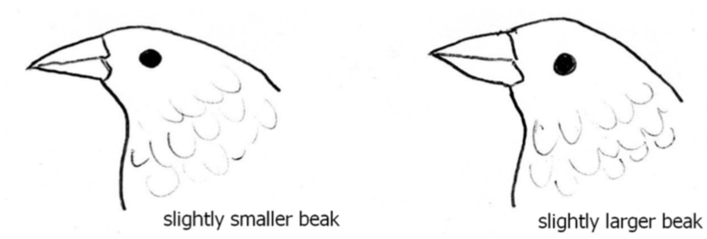
3. Which of these answers is **TRUE** about the evolution of plants and animals? The circles show organisms that share a common ancestor. **Select one answer from the options below.**



4. What would happen to a species of lizards when a new predator appears where the lizards live?

A. Lizards that already have traits that help them to avoid the predator would be more likely to survive and reproduce. The ones that don't would be less likely to survive and reproduce.
B. All lizards would try to develop new traits to avoid the new predator.
C. Some lizards would try to develop new traits to avoid the new predator. The others would die.
D. All the lizards have the same traits because they are from the same species. None of them can survive and reproduce better than anyone else. They would all either die or survive.

5. A species of bird ate many types of seeds. Members of the species had different sized beaks.



The climate changed to be much drier and only much larger seeds were left.

Which answer best explains why almost all of the birds had slighter larger beaks after many generations?

A. The birds with bigger beaks were better at eating the larger seeds. Only birds with big beaks got enough food to survive, reproduce and pass the big beak trait on to the next generation.
B. Birds with smaller beaks had to work really hard to open the bigger seeds. The more they used their beaks the larger they grew. They were then able to get enough food to survive, reproduce and pass on the big beak trait.
C. Birds with smaller beaks grew their beaks so they could open the seeds to get enough food to survive, reproduce and pass on the big beak trait.
D. It was chance that all the birds' beaks got bigger in one generation. They could then get enough food to survive, reproduce and pass on the big beak trait.

6. Could members of a species today look different to individuals of the same species from many generations ago?

A. **YES**. All individuals can change a little in their lifetime and pass on those changes to their offspring.
B. **YES.** Some individuals can change a little in their lifetime and pass on those changes to their offspring.
C. **YES**. Some individuals born with certain traits are more likely to survive and pass on these traits to their offspring.
D. **NO**. Species do not change even after many generations. So individuals of the same species wouldn't look different.

7. Which of these is **NEEDED** for the process of natural selection to occur?

A. Members of the same species must compete with each other.
B. Members of different species must compete with each other.
C. There must be a sudden environmental change.
D. Traits must be inherited from one generation to the next

8. Members of the same species can inherit different traits. What could these differences change? **Select one answer from the table below.**

Answer	Their ability to find food	Their ability to attract a mate
A	✓	✓
B	✓	✗
C	✗	✓
D	✗	✗

9. Which answer is **TRUE** about members of the same species?

A. They may have inherited different traits. These differences may change their chance of surviving and reproducing.
B. They would all have the same inherited traits because they are from the same species. So they all have an equal chance of surviving and reproducing.
C. They all have the same inherited traits. But they pick up different skills and learn new things during their lifetime. Only these new skills change their chance of surviving and reproducing
D. They may have inherited different traits. But these different traits do not change their chance of surviving and reproducing.

10. Which answer best describes what a fossil can be? **Select one answer form the table below.**

Answer	A bone in which the original matter has been replaced by rock	An impression left by a bone in rock
A	✓	✗
B	✗	✓
C	✓	✓
D	✗	✗

11. Which answer is **TRUE** about extinction of species?

A. Very few species have ever become extinct. Most are still alive.
B. There have been extinction events when many species became extinct at about the same time. Apart from these, extinction is very rare.
C. Humans have caused most of the extinctions. Up until recently species rarely became extinct.
D. Many species have become extinct throughout the history of life on earth.

12. A scientist finds some fossils of an extinct species of fish. What could she do by studying the fossils? **Select one answer from the table below.**

Answer	Discover what features the extinct species had	Discover similarities and differences between features of the extinct species and those alive today
A	✓	✓
B	✓	✗
C	✗	✓
D	✗	✗

13. Which answer is **TRUE** about oak trees and lizards? **Select one answer from the table below.**

Answer	There are similarities between oak trees and lizards	There are differences between oak trees and lizards
A	✓	✓
B	✓	✗
C	✗	✓
D	There is no way to tell	

14. Which answer is TRUE about the species that are living on earth today?

A. All species living today have existed since the time life began.
B. Most species living today have existed since the time life began. But a few have appeared more recently.
C. Most species alive today didn't exist at the time life began.
D. There is no way of finding out. You can't tell whether all, most or a few species living today have existed since the time life began.

15. Which answer is TRUE about how environmental conditions have changed since life began on earth?

A. Conditions have remained about the same everywhere on earth. With only minor changes from year to year.
B. Conditions have remained the same in the oceans but have changed on land.
C. Conditions have remained the same except for a few sudden changes in certain places. Such as a meteorite hitting the earth.
D. Conditions have changed dramatically. Some of these changes have happened suddenly and others more gradually.

<u>The adoption story</u>

There was a woman called Mrs. Smith. Mrs. Smith went into hospital and gave birth to a baby girl. Unfortunately, Mrs. Smith died just after having the baby and she never even got to see or hold her baby. Fortunately, there was a really nice lady called Mrs. Jones who was visiting the hospital. Mrs. Jones had always wanted a little girl and saw the baby girl was all alone after her birth mother died. Mrs. Jones adopted the baby girl and took her home immediately to live with her. Mrs. Jones loved the little girl very much and called her "Daughter" and the little girl loved her very much and called her "Mummy". The little girl spent her whole life with Mrs. Jones.

Think about who the little baby girl was born to? Who did she live with and grow up with?

Now think about what the little girl is going to be like when she grows up into a young woman. Fill in the column for the adopted baby girl to show what features will she have when she grows up.

	Feature	**Mrs. Smith (Birth mother)**	**Mrs. Jones (Adoptive mother)**	**Grown up adopted baby girl**
1	**Skin colour**	White	Black	
2	**Hair colour**	Ginger	Brown	
3	**Favourite TV soap**	East Enders	Coronation Street	
4	**Nose shape**	Turned up at end	Straight	
5	**Tattoo**	Butterfly on shoulder	None	
6	**Shape of pancreas (an organ inside the body)**	Flat	Round	
7	**Chin dimple**	Didn't have one	Does have one	
8	**Eating habits**	Vegetarian	Not vegetarian	
9	**Going to church**	Didn't go to church	Goes to church every week	
10	**Freckles**	Had freckles	Doesn't have freckles	

Now explain your answers in as much detail as possible with what you know about variation.

C3 Power Point version of Student Questionnaire

Student Questionnaire

GEVO2each project

Directions – Personal information

- Fill in your personal details on the front of the answer sheet
- Make sure that you include your date of birth, including the year.
- The name of your science teacher is the teacher that normally teaches you for science.

How to do the test

- Your teacher is now going to read out the questions one by one
- There are 15 multiple choice questions
- Listen carefully to each question
- Circle ONE correct answer (A,B,C or D)
- Your teacher will give you time to think and will let you see the question on the board if you need it.
- You MUST circle an answer for all 15 questions.
- A trait is a characteristic e.g. Eye colour.

Question 1

Some organisms, like a human and a chimpanzee (chimp) have many similarities. Others like a lion and a worm have fewer. The circles show organisms that share a common ancestor.

Which answer is TRUE about who shares a common ancestor?

Select one answer from the options below.

A Humans & Chimps Lions Worms

B Humans & Chimps Lions & Worms

C Humans Chimps Lions Worms

D Humans & Chimps & Lions & Worms

Question 2

Which answer is **TRUE** about a species living today and an extinct species?

(A) A living species and an extinct one **can share a common ancestor,** but only if they have **lots of similarities.**

(B) A living species and an extinct one **can share a common ancestor** that lived a long time ago. Even if they have **few similarities.**

(C) A living species **can't share** a common ancestor with an extinct one.

(D) **Only living species can share a common ancestor** with each other. Extinct species can't share a common ancestor with each other.

Question 3

This question is about the evolution of plants and animals. The circles show organisms that share a common ancestor.

Which answer is TRUE about who shares a common ancestor?

Select one answer from the options below.

A All plants & all animals

B All plants All animals

C All plants All animals

D All animals All plants

Question 4

What would happen to a species of lizards when a new predator appears where the lizards live?

(A) Lizards that **already have traits** that help them avoid the predator would be **more likely to survive and reproduce**. The ones that don't would be less likely to survive and reproduce.

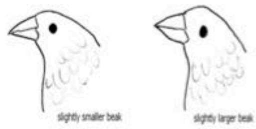
(B) **All** lizards would **try to develop** new traits to avoid the new predator.

(C) **Some** lizards would **try to develop** new traits to avoid the new predator. The others would die.

(D) All the lizards have the **same traits** because they are from the **same species**. All of them have the **same chance** of surviving or dying.

Question 5

A species of bird had different sized beaks and ate many types of seeds.



The climate changed to be much drier and only much larger seeds were left.

Which answer **best explains** why almost all of the birds had slighter larger beaks after many generations?

(A) The birds with **bigger beaks** could eat the big seeds. Only these birds got enough food to survive and **pass on the big beak trait** to their offspring.

(B) Birds with **smaller beaks worked** really hard to open the big seeds. Their beaks got **bigger** because they **used** them. They could then get enough food to survive and pass on the big beak trait.

(C) Birds with **smaller beaks grew** their beaks because they they **wanted** to open the seeds to get enough food to survive and pass on the big beak trait.

(D) It was **chance** that **all the birds' beaks got bigger** in one generation. They could then get enough food to survive and pass on the big beak trait.

Question 6

Could members of a species today look different to individuals of the same species from many generations ago?

(A) **YES**. **All** individuals can change a little in their **lifetime** and pass on those changes to their offspring.

(B) **YES. Some** individuals can change a little in their **lifetime** and pass on those changes to their offspring.

(C) **YES**. Some individuals **born** with certain traits are more likely to survive and pass on these traits to their offspring.

(D) **NO**. Species do not change even after many generations. So individuals of the same species wouldn't look different.

Question 7

Which of these is **NEEDED** for the process of natural selection to occur?

(A) Members of the **same species** must **compete** with each other.

(B) Members of **different species** must **compete** with each other.

(C) There must be a sudden **environmental change**.

(D) **Traits must be inherited** from one generation to the next

Question 8

Members of the same species can inherit different traits.

What could these differences change?

Select one answer from the table below.

Answer	Their ability to find food	Their ability to attract a mate
A	✔	✔
B	✔	✗
C	✗	✔
D	✗	✗

Question 4

What would happen to a species of lizards when a new predator appears where the lizards live?

(A) Lizards that **already have traits** that help them avoid the predator would be **more likely to survive and reproduce**. The ones that don't would be less likely to survive and reproduce.

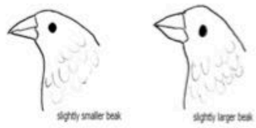
(B) **All** lizards would **try to develop** new traits to avoid the new predator.

(C) **Some** lizards would **try to develop** new traits to avoid the new predator. The others would die.

(D) All the lizards have the **same traits** because they are from the **same species**. All of them have the **same chance** of surviving or dying.

Question 5

A species of bird had different sized beaks and ate many types of seeds.



The climate changed to be much drier and only much larger seeds were left.

Which answer **best explains** why almost all of the birds had slighter larger beaks after many generations?

(A) The birds with **bigger beaks** could eat the big seeds. Only these birds got enough food to survive and **pass on the big beak trait** to their offspring.

(B) Birds with **smaller beaks worked** really hard to open the big seeds. Their beaks got **bigger** because they **used** them. They could then get enough food to survive and pass on the big beak trait.

(C) Birds with **smaller beaks grew** their beaks because they they **wanted** to open the seeds to get enough food to survive and pass on the big beak trait.

(D) It was **chance** that **all the birds' beaks got bigger** in one generation. They could then get enough food to survive and pass on the big beak trait.

Question 6

Could members of a species today look different to individuals of the same species from many generations ago?

(A) **YES**. **All** individuals can change a little in their **lifetime** and pass on those changes to their offspring.

(B) **YES. Some** individuals can change a little in their **lifetime** and pass on those changes to their offspring.

(C) **YES**. Some individuals **born** with certain traits are more likely to survive and pass on these traits to their offspring.

(D) **NO**. Species do not change even after many generations. So individuals of the same species wouldn't look different.

Question 7

Which of these is **NEEDED** for the process of natural selection to occur?

(A) Members of the **same species** must **compete** with each other.

(B) Members of **different species** must **compete** with each other.

(C) There must be a sudden **environmental change**.

(D) **Traits must be inherited** from one generation to the next

Question 8

Members of the same species can inherit different traits.

What could these differences change?

Select one answer from the table below.

Answer	Their ability to find food	Their ability to attract a mate
A	✔	✔
B	✔	✗
C	✗	✔
D	✗	✗

Question 9

Which answer is **TRUE** about members of the same species?

(A) They may have **inherited different traits**. These differences may **change their chance** of surviving and reproducing.

(B) They would **all have the same inherited traits** because they are from the **same species**. So they all have an **equal chance of** surviving and reproducing.

(C) They all have the **same inherited traits**. But they **pick up** different skills and learn new things during their **lifetime**. Only these **new skills change their chance** of surviving and reproducing.

(D) They may have **inherited different traits**. But these different traits **do not change their chance** of surviving and reproducing.

Question 10

Which answer **best describes** what a fossil can be?

Select one answer from the table below.

Answer	A bone in which the original matter has been replaced by rock	An impression left by a bone in rock
A	✔	✗
B	✗	✔
C	✔	✔
D	✗	✗

Question 11

Which answer is **TRUE** about extinction of species?

(A) **Very few species** have ever become **extinct**. Most are still alive.

(B) There have been **extinction events** when **many** species became **extinct** at about the **same time**. Apart from these extinction is very rare.

(C) **Humans** have caused the majority of extinctions. Up until recently species rarely became extinct.

(D) **Many species** have become **extinct throughout the history** of life on earth.

Question 12

A scientist finds some fossils of an extinct species of fish.

What can she do by studying the fossils?

Select one answer from the table below.

Answer	Discover what features the extinct species had	Discover similarities and differences between features of the extinct species and those alive today
A	✔	✔
B	✔	✗
C	✗	✔
D	✗	✗

Question 13

Which answer is **TRUE** about oak trees and lizards?

Select one answer from the table below.

Answer	There are similarities between oak trees and lizards	There are differences between oak trees and lizards
A	✓	✓
B	✓	✗
C	✗	✓

D	There is no way to tell

Question 14

Which answer is **TRUE** about the species that are living on earth today?

(A) **All species living** today have **existed since the time life began.**

(B) **Most** species living today **have existed since the time life began.** But a **few** have appeared more **recently.**

(C) **Most** species alive today **didn't exist** at the time life began.

(D) There is **no way of finding out.** You can't tell whether all, most or a few species living today have existed since the time life began.

Question 15

Which answer is **TRUE** about how environmental conditions have changed since life began on earth?

(A) Conditions have remained about the **same everywhere** on earth. Only **minor changes** from year to year.

(B) Conditions have remained the **same in the oceans** but have **changed on land**.

(C) Conditions have remained **the same except** for a **few sudden changes** in **certain places**. Such as a meteorite hitting the earth.

(D) Conditions have **changed dramatically**. Some of these changes have happened suddenly and others more gradually.

C4 Student answer grid for use with Power Point Questions

Student questionnaire

Instructions:

- Please fill in this page of the booklet.
- This assessment has 2 sections.
- There are 15 multiple-choice questions and then a table about an adoption story to complete.
- Answer all of the questions in the booklet.
- Please do not work with anyone else or share answers
- Don't worry if you don't know an answer.

Your Name: ___________________.

Your Date of Birth (including year)___________________

Are you a Boy or a Girl?____

Name of your School: ___________________________

Name of Science Teacher: ______________________

Date:___________________

- Your answers will be kept strictly confidential during this study.
- Your answers will not affect your science grades at school.
- By completing this questionnaire you will be helping research by the Evolution Education Trust and the University of Bath.

Thank you for your help

THE EVOLUTION EDUCATION TRUST

UNIVERSITY OF BATH

SECTION A Multiple Choice Questions

- Your teacher is now going to read out the questions one by one
- There are 15 questions
- Listen carefully to each question
- Circle ONE correct answer (A, B, C or D) in the table below.
- Your teacher will give you time to think and will let you see the question on the board.
- You MUST circle an answer for all 15 questions.

1	A	B	C	D
2	A	B	C	D
3	A	B	C	D
4	A	B	C	D
5	A	B	C	D
6	A	B	C	D
7	A	B	C	D
8	A	B	C	D
9	A	B	C	D
10	A	B	C	D
11	A	B	C	D
12	A	B	C	D
13	A	B	C	D
14	A	B	C	D
15	A	B	C	D

SECTION B The Adoption Story

There was a woman called Mrs. Smith. Mrs. Smith went into hospital and gave birth to a baby girl. Unfortunately, Mrs. Smith died just after having the baby and she never even got to see or hold her baby. Fortunately, there was a really nice lady called Mrs. Jones who was visiting the hospital. Mrs. Jones had always wanted a little girl and saw the baby girl was all alone after her birth mother died. Mrs. Jones adopted the baby girl and took her home immediately to live with her. Mrs. Jones loved the little girl very much and called her "Daughter" and the little girl loved her very much and called her "Mummy". The little girl spent her whole life with Mrs. Jones.

Think about who the little baby girl was born to? Who did she live with and grow up with?

Now think about what the little girl is going to be like when she grows up into a young woman. Fill in the column for the grown up adopted baby girl. Will she be like Mrs. Smith or Mrs. Jones for each different feature?

Feature	Mrs. Smith (Birth mother)	Mrs. Jones (Adoptive mother)	Grown up adopted baby girl
Skin colour	White	Black	
Hair colour	Ginger	Brown	
Favourite TV soap	EastEnders	Coronation Street	
Nose shape	Turned up at end	Straight	
Tattoo	Butterfly on shoulder	None	
Shape of pancreas (an organ inside the body)	Flat	Round	
Chin dimple	Doesn't have one	Does have one	
Eating habits	Vegetarian	Not vegetarian	
Going to church	Didn't go to church	Goes to church every week	
Freckles	Has freckles	Doesn't have freckles	

Now explain your answers in as much detail as possible with what you know about variation.

Appendix D

D1 Pre-teaching Teacher Questionnaire

Evolution Teacher Questionnaire

This questionnaire is in 3 parts and should take around 20 minutes to complete.
Section A: Assesses your views on natural selection and evolution. Adapted from (MATE) Rutledge and Warden (1999).
Section B: Assesses your evolutionary knowledge. Adapted from (CINS) Anderson et al. (2202).
Section C: Assesses your confidence of teaching evolution.
Please complete this survey without the help of any resources. Please answer all questions. Tests will remain anonymous throughout the study.

Background Information

Name: ______________________ **Sex: (Male or Female)** ________________

Name of school: ____________________ **Date:** __________________

Type of school: Primary / Junior / Middle /Other (please specify) __________________

Have you undertaken formal lessons in Evolution?
Yes No If yes when did you study evolution:

O-LEVEL/GSCE	
A-LEVEL	
DEGREE	
OTHER (please specify)	

How many years have you been teaching? ____________

Do you follow a religion?
Yes No Prefer not to say

What is your highest qualification in Biology?

O-LEVEL/GSCE	
A-LEVEL	
DEGREE	
OTHER (please specify)	

By completing this questionnaire you will be helping research by the Evolution Education Trust and the University of Bath.

Thank you for your help

THE EVOLUTION EDUCATION TRUST

UNIVERSITY OF BATH

<u>Section A</u>   This section aims to determine your views about natural selection and evolution.

For the following items, please indicate (**by ticking one number**) your agreement/disagreement with the given statements using the following scale:

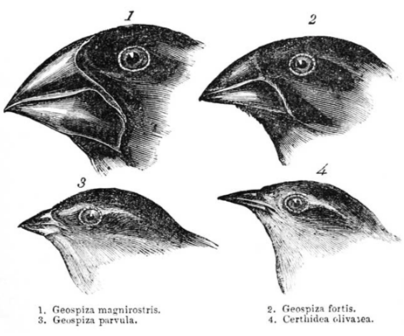
1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly agree

Statement	1	2	3	4	5
1. Organisms existing today are the result of evolutionary processes that have occurred over millions of years					
2. The theory of evolution is incapable of being scientifically tested.					
3. Modern humans are the product of evolutionary processes that have occurred over millions of years					
4. The theory of evolution is based on speculation and not valid scientific observation and testing					
5. Most scientists accept evolutionary theory to be a scientifically valid theory.					
6. The available data are ambiguous (unclear) as to whether evolution actually occurs					
7. The age of the earth is less than 20,000 years					
8. There is a significant body of data that supports evolutionary theory					
9. Organisms exist today in essentially the same form in which they always have					
10. Evolution in not a scientifically valid theory					
11. The age of the earth is at least 4 billion years					
12. Current evolutionary theory is the result of sound scientific research and methodology					
13. Evolutionary theory generates testable predictions with respect to the characteristics of life					
14. The theory of evolution cannot be correct since it disagrees with the Biblical account of creation					
15. Humans exist today in essentially the same form in which they always have.					
16. Evolutionary theory is supported by factual historical and laboratory data					
17. Much of the scientific community doubts if evolution occurs.					
18. The theory of evolution brings meaning to the diverse characteristics and behaviours observed in living forms.					
19. With few exceptions, organisms on earth came into existence at about the same time					
20. Evolution is a scientifically valid theory					

<u>Section B</u> The aim of this section is to assess your evolutionary knowledge.

Your answers to these questions will assess your understanding of the theory of natural selection. Please circle **ONE** answer that best reflects how a biologist would think about each question for each of the three different organisms.

<u>1. Galapagos finches</u>



Scientists have long believed that the 14 species of finches on the Galapagos Islands evolved from a single species of finch that migrated to the islands over five million years ago (Lack, 1940).

Recent DNA analyses support the conclusion that all of the Galapagos finches evolved from the warbler finch (Grant, Grant & Petren, 2001; Petren, Grant & Grant , 1999).

Different species live on different islands. For example, the medium ground finch lives on one island. The large cactus finch lives on another island.

One of the major changes in the finches is in their beak size and shape, as shown in the diagram.

1. What would happen if a breeding pair of finches were placed on an island which had ideal conditions with no predators and unlimited food so that all individuals survived?

   Given enough time,
   (a) the finch population would stay small because birds only have what they need to survive.
   (b) the finch population would double and then stay relatively stable.
   (c) the finch population would increase dramatically
   (d) the finch population would grow slowly and then level off.

2. Finches on the Galapagos Islands require food to eat and water to drink.

   (a) When food and water are scarce, some birds may be unable to obtain what they need to survive.
   (b) When food and water are limited, the finches will find other sources, so there is always enough.
   (c) When food and water are scarce, all finches eat and drink less so that all birds survive.
   (d) There is always plenty of food and water on the Galapagos Islands to meet the finches' requirements.

3. Once a population of finches has lived on a particular island for many years,

   (a) the population continues to grow rapidly.
   (b) the population remains stable, with some fluctuations.
   (c) the population dramatically increases and decreases each year.
   (d) the population will decrease steadily.

4. In the finch population, what are the primary changes that occur gradually over time?

(a) The traits of each finch within a population gradually change.
(b) The proportions of finches having different traits within a population change.
(c) Successful behaviours learned by finches are passed on to offspring.
(d) Mutations occur to meet the needs of the finches as the environment changes.

5. Depending on their beak size and shape, some finches get nectar from flowers, some eat grubs from bark, some eat small seeds, and some eat large nuts. Which statement best describes the interactions among finches and the food supply?

(a) Most of the finches on an island cooperate to find food and share what they find.
(b) Many of the finches on an island fight with one another and the physically strongest ones win.
(c) There is more than enough food to meet all the finches' needs so they don't need to compete for food.
(d) Finches compete primarily with closely related finches that eat the same kinds of food, and some may die from lack of food.

6. How did the different beak types first arise in the Galapagos finches?

(a) The changes in the finches' beak size and shape occurred because of their need to be able to eat different kinds of food.
(b) Changes in the finches' beaks occurred by chance, and when there was a good match between beak structure and the available food, those birds had more offspring.
(c) The changes in the finches' beaks occurred because the environment induced the desired genetic features.
(d) The finches' beaks changed a little bit in shape and size with each successive generation, some getting larger and some getting smaller.

7. What type of variation is passed to the offspring?

   (a) Any behaviours that were learned during a finch's lifetime.
   (b) Only characteristics that were beneficial during a finch's lifetime.
   (c) All characteristics that were genetically determined.
   (d) Any characteristics that were positively influenced by the environment during a finch's lifetime.

8. What caused the populations of birds having different beak shapes and sizes to become distinct species distributed on the various islands?

(a) The finches were quite variable, and those whose features were best suited to the available food supply on each island reproduced most successfully.
(b) All finches are essentially alike and there are not really fourteen different species.
(c) Different foods are available on different islands and for that reason individual finches on each island gradually developed the different beaks they needed.
(d) Different lines of finches developed different beak types because they needed them in order to obtain the available food.

2.Venezuelan Guppies

Guppies are small fish found in streams in Venezuela. Most guppies are brightly coloured, with black, red, blue and iridescent (reflective spots). Males can't be too brightly coloured or they will be seen and eaten by predators. If they are too plain females will choose other males. Natural selection and sexual selection push in opposite directions.

When a guppy population lives in a stream in the absence of predators, the proportion of brightly coloured, flashy males increases. If a few aggressive predators are added to the same stream, the proportion of these flashy males decreases within about five months (3-4 generations).

The effects of predators on guppy colouration have been studied in artificial ponds with mild, aggressive and no predators, as well as similar manipulations of natural stream environments (Endler 1980)

9. A typical natural population of guppies consists of hundreds of guppies. Which statement best describes the guppies of a single species in an isolated population

    (a) The guppies share all of the same characteristics and are identical to each other.
    (b) The guppies share all of the essential characteristics of the species. The minor variations they display don't affect survival.
    (c) All of the guppies are identical on the inside, but have many differences in appearance.
    (d) The guppies share many essential characteristics, but also vary in many features.

10. Fitness is a term often used by biologists to explain the evolutionary success of certain organisms. Which feature would a biologist consider to be most important in determining which guppies were the most "fit"

    (a) Large body size and ability to swim quickly away from predators.
    (b) Excellent ability to compete for food.
    (c) High number of offspring that survived to reproductive age.
    (d) High number of matings with many different females.

11. Assuming ideal conditions with abundant food, space and no predators, what would happen if a pair of guppies were placed in a large pond?

    (a) The guppy population would grow slowly, as guppies would only have the number of babies that are needed to replenish the population.
    (b) The guppy population would grow slowly at first, then would grow rapidly, and thousands of guppies would fill the pond.
    (c) The guppy population would never become very large, because only organisms such as insects and bacteria reproduce in that manner.
    (d) The guppy population would continue to grow slowly over time.

12. Once a population of guppies has been established for a number of years in a real (not ideal) pond with other organisms including predators, what is likely to happen to the population?

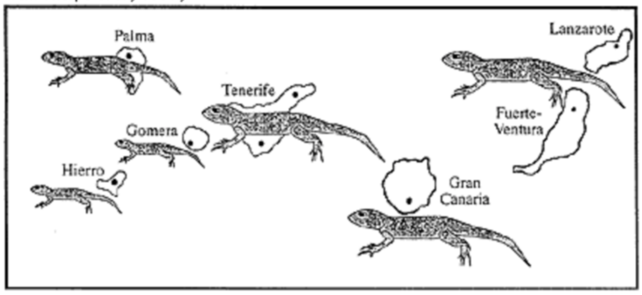
    (a) The guppy population will stay about the same.
    (b) The guppy population will continue to rapidly grow in size.
    (c) The guppy population will gradually decrease until no more guppies are left.
    (d) It is impossible to tell because populations don't follow patterns.

13. In guppy populations, what are the primary changes that gradually occur over time?

    (a) The traits of each individual guppy within a population gradually change.
    (b) The proportions of guppies having different traits within a population change.
    (c) Successful behaviours learned by certain guppies are passed on to offspring.
    (d) Mutations occur to meet the needs of the guppies as the environment changes

<u>3. Canary Island Lizards</u>

Figure 2. The relative sizes of typical lizards from each population are shown. (Redrawn from Thorpe et al., 1994.)



The Canary Islands are seven islands just west of the African continent. The islands gradually became colonized with life: plants, lizards, birds, etc.

Three different species of lizards found on the islands are similar to one species found on the African continent (Thorpe & Brown, 1999).

Because of this, scientists assume that the lizards travelled from Africa to the islands by floating on tree trunks washed out to sea.

14. Lizards eat a variety of insects and plants. Which statement describes the availability of food for lizards on the Canary Islands?

    (a) Finding food is not a problem since food is always in abundant supply.
    (b) Since lizards eat a variety of foods, there is likely to be enough food for all the lizards at all times.
    (c) Lizards can get by on very little food, so the food supply does not matter.
    (d) It is likely that sometimes there is enough food, but at other times there is not enough food for all the lizards.

15. What do you think happens among lizards of a certain species when the food supply is limited?

(a) The lizards compete for food and share what they find.
(b) The lizards fight for the available food and the stronger lizards kill the weaker ones.
(c) Genetic changes that would allow the lizards to eat new food sources are likely to be induced.
(d) The lizards least successful in the competition for food are likely to die of starvation and malnutrition.

16. Populations of lizards are made up of hundreds of individual lizards. Which statement describes how similar they are likely to be to each other?

(a) All lizards in the population are likely to be nearly identical.
(b) All lizards in the population are identical to each other on the outside, but there are differences in the internal organs such as how they digest food.
(c) All lizards in the population share similarities, but there are differences in features like body size and claw length.
(d) All lizards in the population are completely unique and share no features with other lizards.

17. Which statement could describe how traits in lizards pass from one generation of lizards to the next generation?

(a) Lizards that learn to catch a particular type of insect will pass the new ability to their offspring
(b) Lizards that are able to hear, but have no survival advantage because of hearing, will eventually stop passing on the 'hearing' trait.
(c) Lizards with stronger claws that allow them to catch certain insects have offspring whose claws gradually get even stronger during their lifetime.
(d) Lizards with a particular colouration and pattern are likely to pass the same trait on to their offspring.

18. Fitness is a term often used by biologists to explain the evolutionary success of certain organisms. Below are descriptions of four fictional female lizards. Which lizard might a biologist consider to be the 'most fit'?

Feature	Lizard A	Lizard B	Lizard C	Lizard D
Body length (cm)	20	12	10	15
Number of offspring surviving to adulthood	19	28	22	26
Age at death (years)	4	5	4	6
Comments	Very healthy, strong and clever	Mated with many lizards	Dark coloured and very quick	Has largest territory

(a) Lizard A
(b) Lizard B
(c) Lizard C
(d) Lizard D

19. According to the theory of natural selection, where did the variation in body size in the three species of lizards most likely come from?

    (a) The lizards needed to change in order to survive, so beneficial new traits developed.
    (b) The lizards wanted to become different in size, so beneficial new traits gradually appeared in the population.
    (c) Random genetic and sexual recombination both created new variations.
    (d) The island environment caused the genetic changes in the lizards.

20. What would cause one species to change into three species over time?

(a) Groups of lizards encountered different island environments so the lizards needed to become new species with different traits in order to survive.
(b) Groups of lizards must have been geographically isolated from other groups and random genetic changes must have accumulated in these lizard populations over time.
(c) There may be minor variations, but all lizards are essentially alike and all are members of a single species.
(d) In order to survive, different groups of lizards needed to adapt to the different islands, and so all organisms in each group gradually evolved to become a new lizard species.

<u>Section C</u> The aim of this section is to assess your confidence of teaching evolution.

For the following items, please indicate (**by circling one number**) your level of confidence with the given statements, using the following scale:

1	2	3	4	5
Not confident at all	Not confident	Fairly confident	Confident	Really confident

Please circle the number that most closely reflects your view

How confident are you about teaching science in general?

1 2 3 4 5

How confident are you about teaching evolution in general?

1 2 3 4 5

How confident are you about teaching students how to "Recognise that living things have changed over time and that fossils provide information about living things that inhabited the Earth millions of years ago."

1 2 3 4 5

How confident are you about teaching students how to "Recognise that living things produce offspring of the same kind, but normally offspring vary and are not identical to their parents."

1 2 3 4 5

How confident are you about teaching students how to "Identify how animals and plants are adapted to suit their environment in different ways and that adaptation may lead to evolution."

1 2 3 4 5

Do you have any major concerns about the teaching of evolution?

D2 Post-teaching Teaching Questionnaire

Over all feedback on the GEVO2teach evolution topic

The aim of this section is to assess your confidence levels and future needs

1. For the following items, please indicate (**by circling one number**) your level of confidence with the given statements using the following scale:

1	2	3	4	5
Not confident at all	Not confident	Fairly confident	Confident	Really confident

Please circle the number that most closely reflects your view

How confident are you about teaching science in general?

1 2 3 4 5

How confident are you about teaching evolution in general?

1 2 3 4 5

How confident are you about teaching students how to "Recognise that living things have changed over time and that fossils provide information about living things that inhabited the Earth millions of years ago."

1 2 3 4 5

How confident are you about teaching students how to "Recognise that living things produce offspring of the same kind, but normally offspring vary and are not identical to their parents."

1 2 3 4 5

How confident are you about teaching students how to "Identify how animals and plants are adapted to suit their environment in different ways and that adaptation may lead to evolution."

1 2 3 4 5

2. For the following items, please indicate (**by circling one number**) how your level of confidence has changed through teaching the GEVO2teach evolution topic, for the given statements using the following scale

1	2	3	4	5
Significantly less confident	Slightly less confident	No change in confidence level	Slightly more confident	Significantly more confident

Please circle the number that most closely reflects your view for each strand of the NC

Has your confidence level changed for "Recognise that living things have changed over time and that fossils provide information about living things that inhabited the Earth millions of years ago."

1 2 3 4 5

Has your confidence level changed for "Recognise that living things produce offspring of the same kind, but normally offspring vary and are not identical to their parents." 1 2 3 4 5

Has your confidence level changed for "Identify how animals and plants are adapted to suit their environment in different ways and that adaptation may lead to evolution."

1 2 3 4 5

3. Do you have any major concerns about teaching of evolution?

4. Do you think there are any areas of evolution that you need further support with, so that you can teach it effectively?

5. How could the GEVO project help to support you?

Appendix E

Trial school fieldnotes

Observations and feedback from School 1 Scheme of Work 1

lesson 1 3/6/2015, 1.30-3.30pm
n=15 year 6 (one student at dentist)
Mr M teaching the lesson directly from the scheme of work with me observing the class and obtaining feedback.

Questionnaires version 1 full written version (1.30-2.00 pm)

Students completed the questionnaires in silence with no intervention from the class teacher or researcher, except to direct them to answer Q8, 10 and 12 when it was noticed that they were omitted by some of the students.

Observation	Students	Teacher (Mr M)
Completion time (minutes)	20-30	15
Amount of reading	1 too much, 14 OK	OK
Difficulty	All perceived switch story to be easier than multiple choice questions	OK
Understanding the questions	2 students didn't understand the lay out of the table questions (8,10,12). Thought that they were giving information rather than requiring an answer. (Ties in with Galapagos student responses)	No problems but really made him think. "Just have to go with first thoughts rather than over analyse them"
Questions raised during completion	Meaning of the word ancestors. Who their science teacher was?	None

Points raised and discussed

- Mr M Suggested that the reason for the questionnaire taking so long to complete was their SATs tests training. They are taught to read, re-read, answer then check the answer. So may be over thinking the response.
- Plus slower readers struggled with the amount of text even though they reported that it was OK in feedback.
- Mr M stated receiving the Scheme of Work and resources more than made up for having to complete the teacher questionnaire and future feedback sheet.
- Saved him a lot of time and effort.
- Adopted child in class had no problems completing the adoption story and thought it was fine and not offensive or emotionally challenging.

Recommendations and action to be taken

- Students complete initial questionnaire before the lesson so that it doesn't eat into "science time"

- Teacher reads out each individual multiple-choice question and proposed answers in turn and students mark off response on sheet in OMR format. Will save time, alleviate some problems caused by poor literacy skills and prevent over analysis of the questions. Done and used in post teaching questionnaire
- Alter wording of table questions (8,10 and 12) to make sure students known that a response is the expected outcome. Done

<u>Starter activity (2.00- 2.35)</u>

Facilitated by power point presentation - What is Variation?

Points raised and discussed

- Questions and images promoted some detailed and interesting discussion points within the class as demonstrated by the length of time taken. This time frame can be shortened if time is a constraint.
- When choosing family to observe and compare in the power-point presentation I missed that it was a family appearing with a celebrity (Zak Efron). This caused quite a lot of distraction in the class and wasted some time on the discussion on whether it was his family or not.
- Word trait needed further explanation.
- Mr M chose 3 random traits from list and this was enough to ensure no two students were in the same group for everything.

Recommendations and action to be taken

- Check whether there is an adequate explanation of trait in the power-point Done
- Remove and replace picture of family with Zak Efron in it. Done

<u>Main activity (2.40- 3.15)</u>

Based around Looking at variation in the class worksheet. Mr M organized the class into groups of 5 and then let the students organize themselves and collect their own data.

Points raised and discussed

- Students found the hand span exercise easy to follow and where able to estimate, measure and rank the data.
- A few needed extra direction on how to plot the histogram (ie spacing between data columns and scale)
- Students able to collate whole class data independently and found plotting histograms easier and quicker this time.
- Exercise opened up debate on how to define cheek dimples and the subjectivity of observations in science.
- Teacher reported that task worked well, was of an appropriate level, difficulty and had good links to numeracy.

Recommendations and action to be taken

- Check direction about histogram rather than bar chart in Scheme of Work
- Write a reminder to leave a gap on the student sheet near the axes
- Add in a picture of cheek dimples?

Plenary activity (3.15-3.25)

Mr M read the switch story out to the class and then in pairs they discussed and answered the questions on whether the cause of variation in each trait was inherited or environmental.

Points raised and discussed

- Feedback that task was of an appropriate level and that it worked well as a paired activity.
- Students took about 5 to discuss and complete the table.
- Mr M then went over the correct answers before pack up.

Recommendations and action to be taken

- None

Overall teacher feedback

1. Lesson was easy to teach from the Scheme of Work. Well explained and well resourced.
2. Appropriate level.
3. Good activities were chosen – students involved and active.
4. Good links to numeracy and literacy.
5. Activities reinforced scientific methods.
6. Mr M felt that some kind of additional training was required for primary school teachers. They have the classroom management skills needed to teach effectively but lack the experience, knowledge and resources (like these) to teach evolution with confidence.

Observations and feedback from lesson 2

10/6/2015, 1.30-3.30pm
n=16 year 6
Mr M teaching the lesson directly form the Scheme of Work with me observing the class and obtaining feedback.

Starter activity (1.35-2pm)

Facilitated by power point presentation – “what is adaptation”

Points raised and discussed

- Questions and images promoted some detailed and interesting discussion points within the class as demonstrated by the length of time taken. This time frame can be shortened if time is a constraint.
- Students were able to discuss the two examples given in detail using prior knowledge, as they had already encountered them in their environment topic work. They were given 2 minutes to discuss each organism and then went over the new adaptions detailed on the power point presentation.
- Adaptations of cacti new to them.
- Blob game was very successful. Students enjoyed it and were keen to play again. From the class of 16, 2 students survived, this will probably increase in a larger class.

- Game was quite active, noisy but controlled.
- Students' feedback showed that they enjoyed it and understood the principle behind it. It also opened up an good discussion on what natural selection is.
- Teacher was very keen to adapt the blob game to get the students to come up with their own traits and environmental conditions to see who would survive.

Recommendations and action to be taken

- Name cactus as the plant in the power point presentation
- Make sure that have coloured card (green, yellow and red) for blob game.
- Come up with a game pro-forma so students can make up their own blob games.

Main activity (2.00- 3.00pm)

- Based around "Hunting for beads worksheet". Mr M organized the class into groups of 4 and then let the students organize themselves and collect their own data.
- Each table of four was split into pairs and had the same set of coloured beads and paper, to allow sharing of results.
- This activity took longer than expected as we had to amend the method (see below) and start again.

Points raised and discussed

- 3x5 seconds hunting sessions between generations very good time span (range of 0-5 beads in total removed in the time period)
- Need to amend starting number of "prey" from 25 to 10 of each type as we ran out after 2/3 generations as numbers became too large after doubling. Done
- To make it more difficult/fun add that students should hunt using their non-writing hand. Done
- Students found the process of doubling the survivors a bit confusing so add extra columns into the work sheet to make this step more logical. Done
- Add name of beads and coloured paper into the column headers to clarify instructions. Done
- Put name text box on all worksheets to make it easier to allocate work at end of lesson.
- Not all coloured beads got the desired outcomes (only orange and green)

  Summary of results (at end of hunting, after starting with 10 beads of each colour)

Paper colour	Matching beads surviving	Non-matching beads surviving
Blue/red	30	42
White/pink	6	14
Orange/green	22	0
Purple/yellow	10	10

- These results are not conclusive and may be confusing to the students.
- May be caused by both sets of plastic beads standing out equally well on the coloured card which were good matches but not exact colour matches. In addition the 3D beads stood out on the card and were easy to grasp with the forceps.
- Will try a second version using crumpled newspaper and white paper background and matching discs made using a hole punch to see if the degree of camouflaging is better. Done

- Section A the CLOZE exercise worked well, all students understood the task and completed it with the minimum of fuss.
- Section B was less successful as it didn't tie in with the hunting beads activity as well as it should.
- Change the graph to mirror the starting numbers ie. Equal amount of both colour and alter the question from "what type of environmental conditions do you think the prey live in?" to "Explain what happed to the environment the birds lived in over the 10 generations" Done
- Add in section at end of sheet to allow students to reflect on what they learnt from this simulation and tie into natural selection. Done

Plenary activity (3.00-3.25)

Students worked in groups of 3. They discussed the scenario and then completed the HL questions individually.

Points raised and discussed

- Students were able to complete the sheet but it wasn't very exciting.
- Some struggled to get started – maybe they should have been issued with the LL questions.
- Some students thought that the birds could have evolved in 1000 years and so this misconception must be addressed by the answers sheet
- Mr M then went over the correct answers before pack up.

Recommendations and action to be taken

- Look for an alternative plenary activity or amend this one.
- Make sure that teacher summarizes the point of the lesson by adding in a final slide on what natural selection is and why they did these activities. Amend Scheme of Work to highlight the importance of this stage.

Overall teacher feedback

1. Lesson was easy to teach from the Scheme of Work. Well explained and well resourced.
2. Appropriate level.
3. Good activities were chosen – students involved and active and they enjoyed doing them
4. Beads in boxes worked well, easy to give out and keep in one place.
5. Outcomes not as clear as lesson 1 but could be improved.

Observations and feedback from lesson 3

23/6/2015, 1.30-3.30pm
n=16 year 6
Mr M teaching the lesson directly form the scheme of work with me observing the class and obtaining feedback.

Starter activity (1.30-2.10pm) Loo roll time line

- 5 minute introduction to the activities

- Pupils set up time line in the playground using masking tape method (held down every 10 squares and marked directly onto the tape)
- Overall set up and discussion took 30 minutes but did discuss in detail so could limit time spent to 10/15 minutes if set up already.
- This timing has been proven to work in yr6 outreach evolution workshops at Bath University previously.

Points raised and discussed

- We spent about 5 minutes discussing how the time line shows that all living things share a common ancestor, which could bee seen easily by looking back in time.
- This should help to correct common misconception that closely related organisms share a common ancestor but ones with few similarities don't.

Recommendations and action to be taken

- Add in an extra time line card about common ancestors to make sure it is covered and discussed by every teacher at end of time line activity. Done

Main activity (2.10-3.10pm)

Introduction (2.10-2.30pm)

Facilitated by origin of life power point presentation. This activity took longer than expected as the pupils wanted to talk about the subject. Could be contained into a shorter time scale if required.

Points raised and discussed

- New pictures of extinct animals worked well (feed back from Shaw school)
- Evidence on how we know this was the whale common ancestor worked well (feedback from Shaw school)
- Lots of debate on hippo ancestor, they thought it looked more like a rhino

Recommendations and action to be taken

- Change dinosaur picture to a more realistic one ?

Looking at trilobites observation and modeling (2.35-3.10pm)

Work sheet completed in pairs to identify similarities in form by colouring in 3 body sections on diverse range of trilobites. Model making with salt dough worked really well and all able to discuss their trilobite adaptations very clearly.

Points raised and discussed

- Colouring in of body sections took very long time for some pupils whilst others finished very quickly.
- All got the questions on the trilobite common ancestor correct so suggests appropriate level of difficulty despite being unfamiliar with the diagram lay out.

Recommendations and action to be taken

- Suggest that teachers limit colouring in time to focus attention to completing the task

Plenary activity (3.10-3.25)

Students worked independently on the sheet. Teacher then discussed the correct answers in the last 5 minutes.

Points raised and discussed

- Q6,7 and 8 children found the hardest and needed a little extra help thinking through the answers.
- New estimate question ok as suggested by Shaw school.
- Q1-5 no problems at all

Recommendations and action to be taken

- None

Homework Feedback

Mary Anning

Both levels of homework completed by the students. Lower ability students took longer to complete the exercise but this was a reflection of their ability and need for reassurance rather than the difficulty of the task. Students were able to talk about what they had learnt from the comprehension task and appreciated Mary Annings contribution to science.

Factor	Yes	No
Easy to read	13	3
Right amount of text	16	0
Interesting story	16	0
Good illustrations	16	0
Able to understand and answer questions	16	0

Student Reflection

At end of topic get students to write a reflective review on what they have learnt;

- Evolutionary knowledge gained
- Personal skills gained
- Problems encountered

This could be a good way of confirming/validating the increase in assessment scores. Done for school 2

Observations and feedback from School 2 (Scheme of Work 4)

17/6/2015, 9.00-3.15pm
n= 25 year 6 (3 students away)
Mrs K teaching the lesson directly from the scheme of work with me observing the class and obtaining feedback. This class had already been taught the evolution topic but all of the activities trialled were new to the students. This pilot class was taught the whole topic in one day.

Questionnaires

Carried out earlier in the week. Version 2 of the Student questionnaires was read out by the teacher and displayed as separate power point slides.

Points raised and discussed

- Mrs K reported that the students took 55 minutes to complete their questions. The students asked a lot of questions about the assessment items and perhaps she discussed them too much.
- She thought the teacher questionnaire was quite hard and that on reflection maybe her confidence with the topic was better than first reported before teaching the content.
- Mrs K was educated in Scotland and her undergraduate course contained Biology and perhaps she has a higher level of evolutionary knowledge than most other primary school teachers.

Recommendations and action to be taken

- Look at wording of the questionnaires again to see if they can be simplified any more to make them easier to understand?
- Make sure the teachers don't give too much help/explanation when answering the questionnaires.

<u>Lesson 1</u>

<u>Starter activity (9.10-9.40)</u>

Facilitated by power point presentation - What is Variation?

Points raised and discussed

- Questions and images promoted some detailed and interesting discussion points within the class as demonstrated by the length of time taken. This time frame can be shortened if time is a constraint.
- Family photo fine this time. No problem for the students to spot similarities and differences.
- Puppies gave ooh factor again, could be distracting??
- Words trait, inherit, offspring and variation needed further explanation.
- Mrs K chose 3/4 random traits from list and this was enough to ensure no two students were in the same group for everything.

Recommendations and action to be taken

- Check whether there is an adequate explanation of trait and other words in the power-point. Done
- Produce key words cards for display in the classroom. Not done yet
- Maybe take picture of my hens to replace the puppies?
- Add word <u>naturally</u> to table of traits for straight/curly hair. Done

<u>Main activity (9.40- 10.25)</u>

Based around Looking at variation in the class worksheet. Mrs K organized the class into groups of 4 and helped the students collect their own data.

Points raised and discussed

- Students found the hand span exercise easy to follow and where able to estimate, measure and rank the data.
- Students enjoyed the estimation but really wanted to cheat.

- Took measuring very seriously and measured to the nearest mm.
- Most under estimated.
- The class needed extra direction on how to plot the bar chart (ie spacing between data columns and scale), choice of scale and how to write a title. This was not a problem with the activity, just this class.
- Teacher reported that task worked well, was of an appropriate level, difficulty and had good links to numeracy.

Recommendations and action to be taken

- Check direction grid lines are visible on the photocopied sheets.
- Change wording of sheet from experiment to investigation. Done
- Maybe give more confident teachers the option of students designing their own investigations.

Plenary activity (10.50-11.00)

Students read the sheet independently and then worked with a learning partner to answer the questions.

Points raised and discussed

- Feedback that task was of an appropriate level and that it worked well as a paired activity.
- Students took about 5 to discuss and complete the table.
- Mrs K then went over the correct answers.
- All students reported 4 or 5 finger understanding of the lesson which equates to excellent/very good understanding.

Recommendations and action to be taken

- None

Overall teacher feedback

7. Lesson was easy to teach from the Scheme of Work. Well explained and well resourced.
8. Appropriate level.
9. Good activities were chosen – students involved and active.
10. Good links to numeracy and literacy.
11. Activities reinforced scientific methods.

lesson 2

Starter activity (11.05-11.20) (11.20-11.30 blob game)

Facilitated by power point presentation – “what is adaptation?”

Points raised and discussed

- Questions and images promoted some detailed and interesting discussion points within the class as demonstrated by the length of time taken. This time frame can be shortened if time is a constraint.

- Students were able to discuss the two examples given in detail using prior knowledge, as they had already encountered them in their environment topic work. They were given 2 minutes to discuss each organism and then went over the new adaptions detailed on the power point presentation.
- Adaptations of cacti new to them.
- Blob game was very successful. Students enjoyed it and were keen to play again. From the class of 25 only 1 student survived.
- Game was quite active, noisy but controlled.
- Students' feedback showed that they enjoyed it and understood the principle behind it. It also opened up a good discussion on what natural selection is.
- Also corrected alternate conception that organisms can adapt during their life-time, really understood that they needed to be born with a different trait to survive.
- Teacher was very keen to adapt the blob game to get the students to come up with their own traits and environmental conditions to see who would survive.

Recommendations and action to be taken

- Name cactus as the plant in the power point presentation and add a sentence why they have spines not leaves. Done
- Check wording of blob game to <u>stress</u> being born with a trait not developing one. Done

<u>Main activity (11.30- 12pm) (11.50-12.00 writing task)</u>

Based around "What happened to the peppered moth?" power point presentation, simulation and worksheet consolidation sheet. Mrs K read out and discussed each slide with her class. They were unfamiliar with this example but easily understood how natural selection was operating.

Points raised and discussed

- Simulation game very successful on interactive white board as 2 volunteers could just tap the board and everyone could see and offer "advice"
- Results from the simulation were remarkably similar for both backgrounds. From 50/50 ratio of pale to dark moths, got 62% P and 38%D on pale bark and 37%P and 63%D on dark bark.
- H Worksheet consolidation completed by majority of students with minimum of help needed. Small group of 3 students completed L sheet with help from the teacher.

Recommendations and action to be taken

- On dark slide add caused by <u>chance</u> Done
- On magpie slide add in <u>daytime</u> Done

<u>Plenary activity after lunch (1.06-1.15)</u>

Instead of writing on the sheets Mrs K read out the L level scenario. Each question was discussed with their learning partner before sharing with the whole class.

Points raised and discussed

- This method of delivery worked really well and allowed students to talk and correct alternate conceptions about time scales involved with evolution and the process of natural selection.
- Focus was on discussion rather than writing
- Avoided the use of another piece of paper.

•

Recommendations and action to be taken

- Change this activity into a discussion lead by the teacher following the worksheet format. Done
- Make sure that teacher summarizes the point of the lesson by adding in a final slide on what natural selection is and why they did these activities. Done
- Amend Scheme of Work to highlight the importance of this stage. Done
- Check wording of bird scenario to make sure they can't swim/fly away Done

<u>lesson 3 (1.15-3.15)</u>

<u>Starter activity (1.15- 1.45 including set up)</u>

Facilitated by "loo roll time line" which was set up in a long corridor as too windy to do outside. Students helped to set it up and then the class moved along it discussing what happened at each chosen time spot.

Points raised and discussed

- Can be set up during the lesson rather than before and students can help to set it up very successfully.
- Hold loo roll in place every 10 squares with line of masking tape and then can write numbers on tape rather than needing post it notes as well.

Recommendations and action to be taken

- Change wording of teacher instruction sheet to reflect use of masking tape and not post its. Done
- Give idea of timing with and without prior set up. Done

<u>Main activity (1.45 – 2.45) (1.45-1.55 ppt) (2.00-2.45 forelimbs sheet and model making)</u>

Facilitated by power point presentation – "the origin of life" and "looking at forelimbs" worksheet.

Points raised and discussed

- Power point presentation worked well
- Students worked with learning partner on sheet and salt dough modelling.
- Students had no problem identifying and colouring in the different bones despite their teacher's initial reservations that the task was too hard.
- Students able to suggest names for the different animals and were able to justify their choices with sound reasoning.
- Really enjoyed the salt dough exercise and were able to verbalise the link between humans and other mammals to their common ancestor.
- Mrs K finished off the activity by discussing the limbs of other tetrapods (frog and bird) and how this is evidence of sharing the same common ancestor. Students were easily able to pick out similarities and differences and discuss them.

Recommendations and action to be taken

- To ppt add in how we know this was the whale common ancestor i.e. What evidence has been used. Done
- Add in a slide showing other examples of extinct organisms (plants and animals) to show the diversity. Maybe get them to fly in on a timer over one another. Done

- Add in instruction for teachers to show all tetrapod forelimbs and discuss its significance. Done
- Add in instruction to teacher to have extra flour available and not to allow the students to over work the dough as it can get really sticky.

<u>Plenary (2.55 – 3.05)</u>
Facilitated by worksheet– “Common ancestor”

Points raised and discussed
- Students read and discussed the sheet with their learning partners.
- Then answers discussed as a whole class.
- Definition of common ancestor and extinct fine. (Q1 and Q2)
- Students were able to give divergence dates easily, so understood diagram (Q3-5)
- Q6 easy to state but explanation Q7 was harder but manageable.
- Found Q8 harder and needed a bit more teacher direction. But good answers given.
- Also able to estimate when gorillas branched off.
- On initial inspection of the worksheet Mrs K thought this task would be too hard but students more than able to understand the diagram and answer the questions.

Recommendations and action to be taken
- Add in estimation question about gorillas. Done

<u>Overall teacher feedback</u>

6. Lessons were easy to teach from the Scheme of Work.
7. Scheme of Work was well explained and well resourced.
8. Perhaps too many sheets of paper.
9. Give suggestions for greater flexibility of use of resources in Scheme of Work
10. Despite initial reservations about the perceived difficulty of some of the tasks, students were able to complete them and understand them, so appropriate level of difficulty.
11. Good activities were chosen – students involved and active and they enjoyed doing them.
12. Especially enjoyed the blob game, loo roll time line and modelling.
13. Mrs K had looked at the time line before but it was too time consuming and difficult to set up.
14. Students in this class usually only focus on one subject for 2 hours max. They were able to remain focused and enthusiastic all day due to varied nature of lessons and activities.
15. The school has a deputy head teacher who has a National award for science teaching in primary school. Consequently, learning science through investigation is embedded in the ethos of the school. Students often design their own investigations. Would have liked to have been able to design their own looking at variation in the class investigation given some examples to look at. <u>Maybe design an alternative sheet for classes that can confidently design their own investigations</u>.
16. Fossils in year 3 is new to the National Curriculum this year too and so some current year 4/5/6 classes may have missed this aspect if the teachers hadn’t taught it retrospectively. <u>Maybe add an optional extra lesson to cover this aspect.</u>

Homework Feedback
Due to time constraints formal feedback on the homework tasks did not take place, however, Mrs K evaluated them as appropriate for the age and ability of her classes using her expert experience as a primary school teacher.

Student Reflection Done

At end of topic students wrote a reflective review on what they have learnt;

- Evolutionary knowledge gained
- Personal skills gained
- Problems encountered
- Good way of confirming/validating the increase in assessment scores.

Appendix G

G1 Telephone script for initial Primary/Junior school contact

Hello, my name is Dana Buchan and I am a PhD researcher at the University of Bath.

I am trying to get hold of name of *yr6 class teacher/science coordinator here.*

Can I take 5 minutes of your time to ask you to help me conduct a survey for my PhD thesis?

I am completing a PhD on the “most effective way to teach evolution to year 6 students and whether the views/knowledge of their teachers influences this”. I have selected your school at random for inclusion in my study.

If you are willing to participate I will provide a fully resourced and detailed differentiated Scheme of Work including homework, to enrich your teaching of the new KS2 evolution topic. All 4 lessons involve active learning and scientific practical skills. I would also like to assure you that the resources are written in a way as not to conflict with any religious beliefs.

The student survey is 20 questions long and will take about 20 minutes to complete. I would ask you to carry out the survey twice, once before teaching the topic and soon after teaching it to assess the effectiveness of the lessons. The teacher survey is 45 questions long and will take about 20 minutes to complete. All information I gather will be strictly confidential: my interest only concerns statistical relationships among the factors investigated by this project. Your school will have no method of determining respondent’s answers to any of the questions. All returned questionnaires will be kept securely and destroyed after my graduation in 2018.

Please consider carefully whether you would be willing to participate in this study. I literally cannot complete it without your help. The project is intended to help and support primary/junior teachers who may not be science specialists. All lessons and resources have been written and evaluated for use specifically with year 6 students to cover the new NC evolution strands.

Ideally, I would like to send you the Scheme of Work and resources on a CD disk and printed copies of the questionnaires and resources for you to use. If you can tell me the number of students you have then I can ensure that the correct number are copied to save you the expense of photocopying.

I realise that you are very busy but does this sound like something you might be able to help me with?

If the teacher answers “yes” then I will say: Thank you for your help. I send you a package containing everything you need to teach the evolution topic in the mail in the next week. I really appreciate your help!

If the teacher answers “no” then I will say: I understand and appreciate your consideration. I do not want to take up too much of your time, but to make my research more statistically meaningful can I ask you why you don’t want to be part of my survey? (I will leave an open-ended question at this point and record any responses.

G2 Copy of initial contact email/letter sent to Primary/Junior school contact

Date
Dear Colleague

Hello, my name is Dana Buchan and I am a PhD researcher at the University of Bath. I am completing a PhD on the best way to teach evolution to year 6 students and whether the views/knowledge of their teachers' influences this. I have selected your school at random for inclusion in my study.

If you are willing to participate I will provide you with a fully resourced and detailed differentiated Scheme of Work including homework, to enrich your teaching of the new KS2 evolution topic. All 4 lessons involve active learning and scientific practical skills. I would also like to assure you that the resources are written in a way as not to conflict with any religious beliefs.

The student survey is 20 questions long and will take about 20 minutes to complete. I would ask you to carry out the survey twice, once before teaching the topic and soon after teaching it to assess the effectiveness of the lessons. The teacher survey is 45 questions long and will take about 20 minutes to complete. All information I gather will be strictly confidential: my interest only concerns statistical relationships among the factors investigated by this project. Your school will have no method of determining respondent's answers to any of the questions. All returned questionnaires will be kept securely and destroyed after my graduation in 2018

Ideally, I would like to send you the Scheme of Work and resources on CD and printed copies of the questionnaires and resources for you to use. If you can tell me the number of students and year 6 teachers you have then I can ensure that the correct number are copied to save you the expense of photocopying. I would also appreciate your help in distributing these resources and surveys to the appropriate people in your school.

I realize that you are very busy but I would greatly appreciate your help with completing this project. If you have any questions regarding this research, resources or the surveys you can contact me at the university on 01225 385902 or mobile 07549947874. I am also easily accessible via email at edmllb@bath.ac.uk.

Please consider carefully whether you would be willing to participate in this study. I literally cannot complete it without your help. The project is intended to help and support primary/junior teachers who may not be science specialists. All lessons and resources have been written and evaluated for use specifically with year 6 students to cover the new NC evolution strands.

Please contact me if you do not wish to participate in my project either by phone or email. To make my research more statistically significant it would be a great help if you could tell me the reason why you do not want to be included in this project. Once you let me know you do not wish to participate in my study I will not contact you any further.

However, I sincerely hope that you will participate in this project. I think it is important research and I will gladly share my final results with anyone who is interested. At the end of my PhD all resources will be made available on TES and other free access websites. I addition all participating schools will be fully referenced.

Yours truly, Dana Buchan

G3 Thank you letter and instructions for survey distribution and completion.

Date

Dear Colleague

Thank you for agreeing to participate in my PhD project on the best way to teach evolution to year 6 students and whether the views/knowledge of their teachers' influences this. To serve as a reminder I am Dana Buchan a PhD student and ex head of biology in a secondary school. I am asking for your help to participate in my study as part of my PhD thesis at the University of Bath.

Enclosed in this parcel you should find the following items

- A CD containing a 4 lesson Scheme of Work, all resources, home-works, teacher information sheets and mark schemes.
- A printed copy of all resources for your reference and to make photocopying easier. (If you need any further photocopying please let me know and I will arrange this for you)
- A plain language statement for students explaining the aims of the study what their participation involves.
- The correct number of student questionnaires for 2 separate assessments (once before teaching and then immediately after teaching the topic using my Scheme of Work and resources). These should take 15-20 minutes to complete.
- The correct number of teacher questionnaires. These should take 20 minutes to complete.
- A self-addressed and stamped envelope for the return of the questionnaires (unless I have agreed to collect them by hand)

Please teach the evolution topic following the provided Scheme of Work as closely as possible. This will allow me to make valid comparisons with the data I collect from other participating schools. Furthermore, I would like to reiterate that all information I gather will be strictly confidential. My interest only concerns statistical relationships among the factors investigated by this project. Your school will have no method of determining respondent's answers to any of the questions. All returned questionnaires will be kept securely and destroyed after my graduation in 2018.

I would like to thank you and your pupils, once again for your participating in my project. I literally cannot complete my research without your help. Thank you! Once again if you do have any questions regarding this research I can be contacted at the university on 01225 385902 or 0754974874. I am also easily accessible via email at edmllb@bath.ac.uk.

Yours sincerely,

Dana Buchan

G4 Formal letter sent to the school participant

To whom it may concern

I am Dana Buchan, a first year PhD student in Biology and Biochemistry at the University of Bath. I am conducting a research project entitled *"What is the most effective way of teaching evolution to primary school children?"* This project aims to develop appropriate learning resources and establish students' perceptions and understanding about natural selection and evolution by using a questionnaire. It targets year 6 students as this topic has been added to the National curriculum for the first time (as of September 2014). I therefore would like to ask for your permission to gain access to the students in your school.

I can confirm that the name of your school, its identification, as well as the name of student participants will be kept confidentially and anonymously in my written reports. All returned questionnaires will be kept securely and destroyed after my graduation in 2018.

Your cooperation will be greatly appreciated.

Sincerely,

Dana Buchan
GEVO2teach Project
Postgraduate Researcher
Department of Biology and Biochemistry
University of Bath
BA2 7AY
tel: 01225 385902
email: edmllb@bath.ac.uk

Appendix H

Classroom observations

Observations and feedback from School 1 Scheme of Work 4 lesson 1 Variation

2/5/2017, 1.30-3.30pm
n=17 year 6
Mrs C teaching the lesson directly from the scheme of work with me observing the class and obtaining feedback.

Feedback about Questionnaires carried out that morning

1. Mrs C was surprised by how the children completed the initial questionnaire: students whom she thought would struggle just got on with it whilst 'more able' were very doubtful about their answers and needed more reassurance.
2. Children were surprised that there were more than 1 mass extinction as thought there was only one when the dinosaurs were wiped out.

Starter activity (1.30- 2.10) Power point and discussion

Power point presentation "What is Variation?"

Points raised and discussed

- Students able to discuss variation quite easily suggesting a wide range of phenotypic similarities and differences in humans, dogs and plants.
- Discussion of the table of genetic/environmental went well as able to reason about the different traits. Lots of debate about heart position and whether the language you speak is inherited.
- Continuum exercise created movement and lots of discussion about identical and non-identical twins.
- One identical twin could tongue roll and one couldn't.

Main activity (2.10- 3.00) "Looking at Variation in the class" worksheet

.Mrs C organized the class into groups of 4 and then let the students organize themselves and collect their own data. Carried out activities 2 and 3 first when class together and then onto group data collection.

Points raised and discussed

- Students found the hand span exercise easy to follow and where able to estimate, measure and rank the data. Seemed to enjoy competitive nature of the task.
- A few needed extra direction on how to plot the histogram (ie spacing between data columns, scale, using a pencil and ruler). All able to plot points and scale axes to suit range of values.
- Exercise opened up debate on how to define cheek dimples and the subjectivity of observations in science.
- Teacher reported that task worked well, was of an appropriate level, difficulty and had good links to numeracy.

Plenary activity (3.00-3.20) "Switch story" worksheet

Students read the story out to the class and then in pairs they discussed and answered the questions on whether the cause of variation in each trait was inherited or environmental.

Points raised and discussed

- Feedback that task was of an appropriate level and that it worked well as a paired activity. Some found it a bit confusing but after discussion realized what they had done wrong.
- Students took about 5 to discuss and complete the table.
- Mrs C then went over the correct answers and discussed any mistakes.
- Some students noticed that information was missing about the mother's and that this could affect what the boys look like.

Observations and feedback from School 1Scheme of Work 4 lesson 2 Natural Selection

8/5/2017, 1.30-3.30pm
n=17 year 6
Mrs C teaching the lesson directly from the scheme of work with me observing the class and obtaining feedback.

Starter activity (1.35-1.55pm) "What is Adaptation" Power point and discussion

Points raised and discussed

- Discussion of inherited and environmental variation to recap previous lesson carried out before power point.
- Examples discussed in pairs and then fed back to whole class.
- Blob game was very successful. Students enjoyed it and were keen to play again. From the class of 17, 1 student survived.
- Game was quite active, noisy but controlled.
-

Main activity (1.55- 2.40pm) Natural selection in the Peppered moth

- First part "What happened to the Peppered moths?" power point (1.55-2.10)
- Mrs C went through power point slides allowing time for pairs to discuss the questions and make predictions about what they thought would happen to the moths.
- All were able to predict the results of differential predation and improved survival of mimetic colouration before simulation game played.
- Worksheet then completed individually and after discussion with a partner, (2.10-2.40).
- Meaning of terms prey and predator recapped before activity.
- Students completed blank side first as easier to do freestyle to get ideas down and then CLOZE exercise with more specialist vocab.
- A couple of students read out their own answers and then CLOZE exercise answers gone over.

Points raised and discussed

- First set of hunting simulation results were not as expected.
- Promoted discussion about scientific methodology and the need to repeat experiments.

Plenary activity (2.40-2.50) Case study worksheet

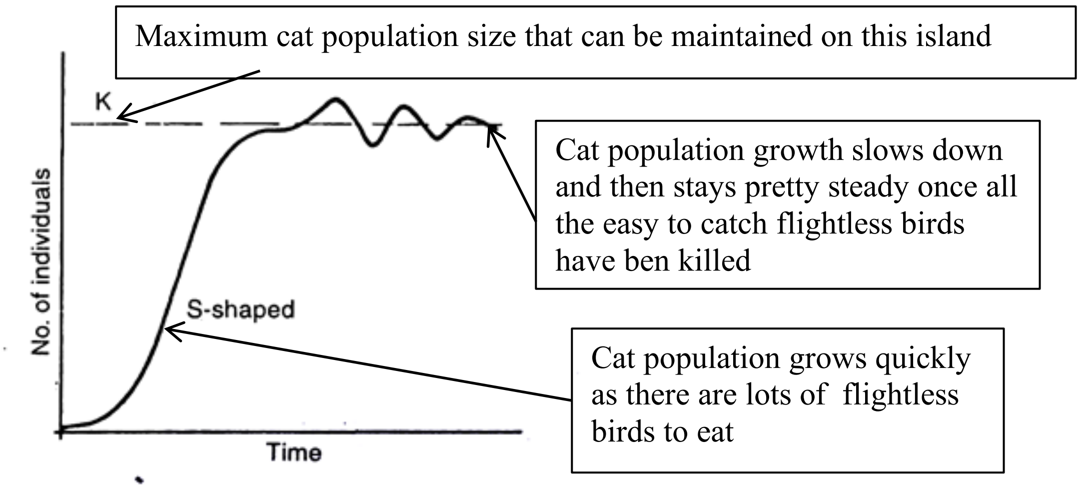
Students worked in groups of 2. They discussed the scenario and then completed the sheet individually.

Points raised and discussed

- Some students asked whether they should include the number of cats as well leading to a discussion of population growth in a confined habitat with a finite food supply.

Recommendations and action to be taken

- Add comment on size of cat population and how it would alter over time on teacher answers



- The session finished a bit early so the simulation game was repeated several times to obtain an average value for each habitat and scientific method discussed.
- Also led onto a discussion on the causes and consequences of albinism in different animals.

Observations and feedback from School 1 Scheme of Work4 lesson 3 Geological time

15/5/2017, 1.30-3.30pm
n=16 year 6 (One student absent due to illness)
Mrs C teaching the lesson directly from the scheme of work with me observing the class and obtaining feedback.

Starter activity (1.30-2.00pm) Video and discussion

- It was a rainy day so start of toilet roll had to be delayed to allow for hall to be cleared after lunch.

- Facilitated by extending the discussion on matters arising from the "Stated Clearly" video

Main activity Toilet roll time line (2.00-2.40pm) and discussion of mass extinctions (2.40-3.00)

- Time line set up in hall and had to double back on itself but still quite manageable.
- Students sorted themselves in time order and then stood where they thought the event occurred
- Students spread out quite equally along the line
- Mrs C then walked everyone along the line correcting errors along the way and promoting the significance of the common ancestor of all living things.
- Mass extinction teacher sheet discussed

Plenary (3-3.20) Spiral time lines

- Nice calm activity after moving about during time line activity

Observations and feedback from School 1 Scheme of Work4 lesson 4 Homology and common ancestry

22/5/2017, 1.30-3.30pm
n=17 year 6
Mrs C teaching the lesson directly from the scheme of work with me observing the class and obtaining feedback.

Starter activity (1.35-2.00pm) Power point and discussion

"Origin of life" power point presentation slides were discussed and questions answered.

Points raised and discussed

- Students interested in marsupials and their distribution
- Also interested in lactose tolerance
- Really seemed to grasp the ladder/tree slide and how they were different. Ladder one looked like higher organisms had evolved from lower ones further down.

Main activity (2.00-2.50pm) Looking at forelimbs and modeling

Work sheet completed in pairs to identify similarities in form by colouring in different forelimb bones on diverse range of mammals. Model making of human arm with salt dough worked really well and all able to model without using arm template by transferring knowledge from monkey limb.

Points raised and discussed

- Concentration levels were good when colouring and lots of discussion about the task eg "this must be for swimming".
- Interested in the range of other tetrapods with the same bone arrangements and linked back to toilet roll time line.

Recommendations and action to be taken

- Give options for the functions on the sheet for the 6 limbs (walking, digging, swimming, flying, running and grasping)
- Check all answers for the common ancestry questions are given

Plenary activity (2.55-3.10) Phylogenetic trees
Students worked independently on the sheet. Teacher then discussed the correct answers in the last 5 minutes.

Points raised and discussed

- No problem with any of the questions, no one asked for help to interpret the cladogram.
- Seemed to find the sheet quite easy

Recommendations and action to be taken

- Make sure answers to all the questions are included on the teacher sheet as seems to be discrepancy (Q6 missing?)
- Maybe do slide of all primates to show diverse range and relationships???

Appendix I

Plain English statement

UNIVERSITY OF BATH

The Evolution Education Trust

Plain Language Statement for students

My name is Dana Buchan. I am a first year PhD student in Biology and Biochemistry at the University of Bath. I would like to invite you to participate in my research project entitled *“What is the most effective way of teaching evolution to primary school children?”* This project is under the supervision of Professor L Hurst

Invitation paragraph

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What are the purposes of the study?

This study aims to investigate (a) The development and assessment of the best scheme of work and teaching resources for the year 6 NC evolution topic (b) Assessment of pupil understanding of evolutionary biology and its evidence, (c) Pupil learning approaches to evolution, and (d) the assessment of the influence of teacher understanding and acceptance of evolution.

Why have I been chosen?

You are being approached because you are a year 6 pupil in a participating primary, junior or middle school.

Do I have to take part?

No. Your participation is voluntary. Although you decide to take part in this research, you are able to withdraw your participation at any time.

What will happen to me if I take part?

If you agree to take part in this research, you will be taught the evolution topic using resources that I have written. You will be given a questionnaire to assess your understanding of natural selection and evolution, before you are taught the topic, just after it and possible 6 months afterwards. Each questionnaire will take approximately 30 minutes to complete.

Will my taking part in this study be kept confidential?

Yes, of course. You will put your names on the questionnaires but when I analyse them I will give you a special code to identify you. Your name will not appear in my study. In addition, the returned questionnaires will be kept in locked filing cabinet at the University of Bath. At the end of my study, all of returned questionnaires will be shredded.

What will happen to the results of the research study?

The results will be analysed by statistical tools as well as qualitatively. The research data will be typed up as a PhD dissertation, journal articles, conference papers and other academic purposes, which you can access (if required).

Who is organising and funding the research?

This research is organised by me, Dana Buchan, through the University of Bath, funded by the Evolution Education Trust. It is under the supervision of Professor L Hurst, the main supervisor and Dr M Hejmadi my second supervisor.

Who has reviewed the study?

Apart from the names of the people mentioned above, this research has been reviewed and approved by ethics committees of the University of Bath.

Contact for Further Information

If you have any other questions about the research project, you can contact me (Mrs Dana Buchan) directly by phone on [redacted] or by email on [redacted].

Appendix J

J1 Student permission forms

Dear Parent/Guardian,

I am pleased to inform you that your son/daughter has been invited to participate in a focus group for the University of Bath's GEVO2teach Primary Evolution Project. Please read this information sheet and complete the attached permission form if you are happy for your son/daughter to participate.

What the project is about

The GEVO2teach Project is an exciting new research initiative that aims to improve the teaching of evolution in primary schools. This study is under the supervision of Professor Laurence D Hurst from the Department of Biology and Biochemistry at the University of Bath. This research is funded by the Evolution Education Trust.

What the focus group is about

The purpose of this focus group is to explore young peoples' views and knowledge of natural selection and evolution. These are topics that your son/daughter has recently studied as part of their new Key Stage 2 National Curriculum using resources developed by this project.

What your son/daughter will be asked to do

Your son/daughter will be asked to participate in a focus group. This will take the form of a relatively informal small discussion group with approximately five other pupils from their science class. Each pupil will be asked opinion-based questions related to what they have learnt in their science classes. They will also be invited to respond to comments other members of the group make. The session will last no longer than one hour and will take place during your son/daughter's normal science lesson, in their school, on a date specified by their teacher. A member of school staff will remain within sight of the group for the duration of the session.

Benefits and risks

Findings from this research may help improve how evolution is taught in primary schools, which could benefit current and future school pupils. This study may prove an interesting experience for pupils to gain insight into biology and education research and will also be an opportunity for pupils to reflect on their learning. No risks greater than those experienced in ordinary conversation are anticipated. Everyone involved in the focus group will be asked to respect the privacy of the other group members.

Taking part is entirely voluntary

Your decision as to whether or not your son/daughter can participate in this focus group will have no impact on their school education or on any current or future relationship with the University of Bath. Your son/daughter will be free to withdraw from the study at any time.

All responses are confidential

The focus group will be recorded using a recording device so that it can be transcribed later; no one else will hear this and the recording will be destroyed once transcribed. Your son/daughter will never be identified by name in the transcription of the focus group or in any reports published as part of this research. All data will be kept strictly confidential: all materials will be stored in a secure location. Data collected will only be seen by members of the GEVO2teach research team.

If you have any queries, please speak to your son/daughter's science/class teacher or contact me directly by phone or email.

If you are **willing** for your son/daughter to be involved in this research, please complete the attached form and return it to their science/class teacher.

I thank you in advance for your support.

Yours faithfully,

Dana Buchan

Dana Buchan
GEVO2teach Primary Project Postgraduate Researcher
Department of Biology and Biochemistry
University of Bath
BA2 7AY

tel: [redacted]
email: l.buchan@bath.ac.uk

Student Focus Group Permission Form

Statement of consent:

I have read and understood the attached information.

By signing below, I give permission for my son/daughter to participate in a focus group for the GEVO2teach Primary Project.

Pupil Name ____

Science/Class Teacher ____

Parent/Guardian Name ____

Signature ____

Date ____

~ Thank you ~

All responses will be kept strictly confidential

J2 Teacher permission forms

<u>GEVO2teach Teacher Focus Group Information Sheet</u>

This focus group is part of my PhD project that is looking at how best to teach Evolution to year 6 students.

- This focus group discussion should last between 45-60 minutes.

- The focus group will be recorded, and then transcribed so that I have a record of the discussion.

- When the focus group's discussion is transcribed I will change your name and any references that you make to specific people and places to ensure that no-one can be identified from our data.

- Nothing you say in this focus group will be heard by anyone else in your school or at the University. All responses are confidential and I ask that all focus group participants respect the privacy and views of other group members.

- If you agree to take part in this focus group, but feel at any stage that you would like to stop, you are free to do so at any time, without giving any reason.

If you have any questions about this project, then feel free to contact me:

Dana Buchan
GEVO2teach Project Postgraduate Researcher
Department of Biology and Biochemistry
University of Bath
BA2 7AY

tel: [redacted]
email: l.buchan@bath.ac.uk

(Participant keeps this section and I have a copy)

Teacher Focus Group Consent Form

Statement of consent:

I have read and understood the information sheet for the GEVO2teach research project and I have had the opportunity to ask any questions I have about the research.

I agree to participate in a focus group for the GEVO2tech project that will be audio-recorded and transcribed. Transcripts will be used only for the purposes of this research project, for as long as this research is being undertaken.

The researchers will not use my personal data for any other purpose or disclose it to any third parties. The information I provide, in the form of my comments, will be anonymised (e.g. my name will be removed and replaced with a number). I agree to parts of what I say being used anomalously by the researchers in publications and presentations.

Participant Name ____

Participant Signature ____

Researcher Name ____

Researcher Signature ____

Date ____

(Two copies required: one to be kept by the interviewee, one to be kept by the researcher)

~ Thank you ~

All responses will be kept strictly confidential

J3 Teacher Focus group questions

Teacher focus group interview questions

Setting the scene and ground rules
Thank you for participating in my PhD project on evolution. My name is Dana Buchan and I am a researcher at the University of Bath. I was a secondary school teacher for 24 years but decided I wanted a new challenge. I was hoping that in the next hour or so we could discuss the evolution and inheritance resources you used. Anything you say will be kept strictly confidential.

Individual introductions
Please can you fill in the name cards so that I can reference any contribution you make in the transcription. Please just say your name as we go around the table.

Initial discussion
1. Were you able to follow and use the Scheme of Work easily?

Discussion
2(a). Which parts/aspects of the teaching intervention package did your students enjoy the most?
2(b). Why do you think this was?
3(a). Did you think the resources helped to correct any misconceptions?
3(b). Which ones?
4. Did you learn anything from using the teaching resources?
5. Did anything surprise you?
6. What is the first word/phrase you think of when I say evolution?
6(a). What did you think of the teaching resources?
6(b). How could they have been improved?
7(a). How would you have approached this topic if you hadn't had access to the project's resources?
7(b). Would you have carried out the practical activities?

Ending discussion
And now the final questions
8. Would you use any of the teaching resources again?
9. Is there anything else before we finish?

Thank you very much for helping me. This session has been really useful for me and it will help me assess the effectiveness of the 4 different Scheme of Work desi designed for the project.